Année : **2010**

THESE

présentée à

L'UFR des Sciences et Techniques
de l'Université de Franche-Comté

pour obtenir le

GRADE DE DOCTEUR DE L'UNIVERSITE DE FRANCHE-COMTE

en Automatique
(Ecole Doctorale Sciences Physiques pour l'Ingénieur et Microtechniques)

Titre du Mémoire

par

Adina Eunice TUŢAC

Soutenue le 22 Octobre 2010 devant la Commission d'examen :

Rapporteurs :
Christian Roux Professeur, U650 INSERM Telecom, Bretagne, France
Henning Müller Professeur, University of Applied Sciences Western Switzerland, Suisse

Examinateurs :
Jacques Klossa Directeur, TRIBVN, France
Nicolas Lomenie Maître de conférences, Université Paris Descartes, France
Nicolae Robu Professeur, Université "Politehnica" de Timişoara, Romania
Noureddine Zerhouni, Professeur des Universités, E.N.S.M.M de Besançon

Directeur de thèse :
Vladimir-Ioan Creţu Professeur, Université "Politehnica"de Timişoara, Romania
Daniel Racoceanu Maître de conférences HDR, CNRS-IPAL UMI 2955, Singapour

Acknowledgments

To complete a PhD thesis project is very difficult by working alone. This is a fact that I experienced during all these years since I started and all the way through the end. Therefore, I desire to thank all the people who where involved or contributed in one way or another to this challenging pursuit.
I would like to express my gratitude to my advisors A. Prof Daniel Racoceanu HDR from Université de Franche-Comté, Besançon and head of IPAL research lab from Singapore and Prof. Vladimir Ioan Cretu from "Politehnica" University of Timisoara for their valuable jointly supervision, experience and professionalism generously displayed throughout our collaboration. This thesis would have not come to life if it were not for their vision and collaborative efforts to offer me the possibility of three research internships in Singapore, at Image and Pervasive Access Lab, an A-STAR Institute for Infocomm Research -I2R Singapore and Centre National de la Recherche Scientifique -CNRS France joint laboratory. I would also want to thank to all my reviewers, Prof. Henning Müller, Prof. Christian Roux, Prof. Noureddine Zerhouni, Prof. Gheorghe Mihalas and Prof. Nicolae Robu for their valuable evaluation of the thesis.

Special thanks to Dr. Joo Hwee Lim from Institute of Infocomm Research, the Singaporean co-director of IPAL and to A. Prof Wee Kheng Leow, from National University of Singapore (NUS), deputy director of the same lab for their support, benevolence in setting the internship stages and their valuable advices. I am also very thankful to all the members of IPAL team and other researchers; I just recall a few: Amel Denappe, Dr. Ludovic Roux, Prof. Nicolas Lomenie, Dr. Mounir Mokhtari, Dr. Chao Hui Huang , Xiong Wei, Jean-Romain Dalle, Li Hao, Antoine Veillard, who encouraged me and help me to enjoy this „slice of life". I remember our lunch times when sharing scientific ideas -mixed with humor- opened a new perspective for me on better understanding a scientific community.

Special acknowledgment goes to M.D. Thomas Putti from the National University Hospital of Singapore (NUH) who provided me all the knowledge about the breast cancer grading and all the slides. His interest and patience along with encouraging discussions on results (even not always good) inspired me not to give up.

I would like to thank to all the co-authors of the papers published at conferences, especially to Dr.Jacques Klossa for his vision, his willingness to collaborate with us and for allowing us to use his platform in our virtual microscope project. I would like to thank to my students, Anca Brandibur and Emina Enikö Szöcs for their interest in collaborating with me at this thesis project.

My gratitude also extends toward Prof Mihai Micea and Claudia Micea for their help both from scientific and administrative standpoints. To Mrs. Agnes Stepanian, Mrs. Cristina Bunea, Mrs. Andreea Lazarescu, Mrs. Lidia Jebelean, Mrs. Simona Damian from the Politehnica University Timisoara Rectorate who gave me a hand when dealing with deadlines of projects and administrative papers of the thesis.

My heartfelt gratitude goes to my dearest friends, Prof. Radu Marinescu and Dr.Cristina Marinescu for their immense help, pertinent advices and encouragement especially in the most difficult moments of this journey. Their support had tremendous value for me. To Mrs. Maria Stolojescu (or Miki as we know her) for her encouragements and delicious morning coffees when preparing for a long day work. To Mr.Seng Kee Koh and his wonderful family that opened their hearts and home when working at the thesis abroad. I cannot but recall my dear friend Thea Skillicorn and her family from Singapore who was there for me in moments of need and also for reviving the good-taste-long-missed Romanian food when not even expecting.

I am deeply thankful to my special friend Roxana Teodorescu with whom I shared the wonderful and also hard times of working on the thesis. Her support cannot be expressed in enough words. To Ildiko Tatai, Mihai Onita, Marlene Daneti, to my colleagues from my office, to all my friends from Romania and Singapore who encouraged me and gave me a hand when needed.

How can I not express my gratitude to my family who stood beside me all this time? To my parents, sister and brothers: your financial and spiritual support has everlasting value. And especially to my husband for his love, encouragements and patience throughout this long way.
And more than to any one else, to God in whom are hidden all the treasures of wisdom and knowledge.

Timişoara, July 2010 Adina Eunice Tuţac

Contents

List of Figures

List of Tables

List of Acronyms

$\mathcal{ALC}$	Attribute Language with Complements
BCG	Breast Cancer grading
BFO	Basic Formal Ontology
CAD	Computer-Aided Diagnosis
CBIR	Content-Based Image Retrieval
CBMIR	Content-Based Medical Image Retrieval
cbPACS	content-based Picture Achieving and Communication System
CBR	Case-Base Reasoning
CV	Computer Vision
DICOM	Digital Imaging and Communications in Medicine
DCIS	Ductal Carcinoma in Situ
DDSM	Digital Database of Screening Mammography
DL	Description Logics
DRD	Diabetic Retinopathy Database
FCA	Formal Concept Analysis
FD	Face Database
FMA	Foundational Model of Anatomy
FOL	First-Order Logic
GTF	Generic Translator Framework
H&E	Hematoxylin & Eozin
HPF	High Power Field
IDC	Invasive ductal carcinoma
IDEM	Images and Diagnosis from Example in Medicine
IRMA	Image Retrieval in Medical Applications
KBR	Knowledge-Based Reasoning
KBS	Knowledge-Based Systems
KR	Knowledge Representation
LCIS	Lobular Carcinoma in Situ

MBR	Model-Based Reasoning
MC	Mitosis Count
MCBR	Medical Case-Based Reasoning
MedGift	Medical GNU Image Finding Tool
MeSH	Medical Subject Heading
MICO	Cognitive Microscope
MK	Medical Knowledge
NCI	National Cancer Institute
NIH	National Institute of Health
NGS	Nottingham Grading System
NNF	Negation Normal Form
NPS	Nuclear Pleomorphism Score
NSD	Non-scale dependency
OBO	Open Biomedical Ontologies
OWA	Open World Assumption
OWL	Ontology Web Language
OWLViz	Ontology Web Language Visualization
PACS	Picture Achieving and Communication System
QBSE	Query By Semantic Example
QBVE	Query By Visual Example
RBR	Rule-Based Reasoning
RDF	Resource Description Framework
RF	Relevance Feedback
RST	Rough Set Theory
ROI	Region of Interest
RuleML	Rule Markup Language
SBML	System Biology Markup Language
$\mathcal{SHIF}$ (D)	ALC with role transitivity, role hierarchy, inverse role, functional role, data types
$\mathcal{SHOIN}$ (D)	ALC with role transitivity, role hierarchy, nominals, inverse roles, number restrictions and data types
SNOMED-CT	Systematized Nomenclature of Medicine- Clinical Terms
SWRL	Semantic Web Rule Language
SPARQL	Protocol And RDF Query Language
SQWRL	Semantic Query Web Rule Language
TISBR	Translational Incremental Similarity Based Reasoning
TF/TFS	Tubule Formation/Tubule Formation Score
UMLS	Unified Medical Language System
VPH	Virtual Physiological Human
VisTex	Vision Texture Database

WHO	World Health Organization
WSI	Whole Slide Imaging
W3C	World Wide Web Consortium

1. Introduction

This thesis addresses ontology-driven prognosis assistance using knowledge representation and reasoning for very large microscopic medical images. This has been done in the context of Whole Slide Imaging exploration for breast cancer grading and prognosis.

1.1. Thesis Context

Domain knowledge representation is an essential component of perceptive and cognitive systems. The formal representation aroused as part of Artificial intelligence field, stirred by the challenge to answer the question: how to formally *think*, what means to use in order to capture the domain of discourse that is *perceived*. Various perspectives on viewing knowledge representation and its characteristics generated arguments over the years. In the paper of [Davis et al., 1993] the need to go back to set the theoretical foundation is highly emphasized and hence the knowledge representation is built on five key pillars. Furthermore, it became evident that knowledge representation and reasoning are "inextricably intertwined" [Davis et al., 1993]. Theories of representation of a wide range of domains developed since then.

One particular direction that is of interest for us is the spatial representation due to its connection with cognition and perception [Aiello, 2002]. Apart from the plethora of approaches in theory and practical implementation [Aiello, 2002], [Cohn and Renz, 2008], the spatial representation and reasoning follows two major paradigms: quantitative and qualitative. The issue of adopting one or the other is highly dependent on the nature and spatial relations of the domain to represent. In a market-context is appropriate to go for a quantitative representation unlike a context that fits the problem of the "aquarium metaphor" [Freksa, 1991]. The situation is illustrated by one observer that wants to describe one particular fish from an aquarium to another observer. However, in this situation, the perception and thus the representation are limited:

- none of the observers can quantitatively locate and position the fish; in a quantitative approach *the positions (not relative positions) of all the objects have to be known, regardless of whether we need them or not*
- not all exact data is available, perceivable features are limited; in a quantitative approach when *a value is not known exactly it has to be either ignored or assigned*
- fuzzy knowledge (the aquarium, the water movement itself may prevent the observers to clearly identify the fish)
- lack of adequacy in language representation, the observers might combine natural language descriptors, the perspectives from where each observer perceives and describes the world of the aquarium might be different, etc.

Although this problem is not universal, it is crucial to identify what type of representation the domain requires, such that falling into the extreme of using quantitative values to express even qualitative facts is avoided [Brageul and Guesgen, 2007].

To this end, this dissertation follows the qualitative representation approach due to its interest in medical applications. This interest relies in the first place on the fact that medical applications are a context where not all data is available when giving a diagnosis or prognosis or these assessments depend heavily on the *perception, the interpretation of facts* of each individual medical expert. In the second place, images (which contain spatial information) play a crucial role in the process of diagnosis or prognosis.

One such example is the domain of breast cancer grading (BCG) which is nowadays considered the key assessment tool in prognosis of modern pathology practice [Frkovic-Grazio and Bracko, 2002], [Steichen et al., 2006]. The recent experiences reveal that the manual histological grading procedure is highly influenced by the competencies and stamina of each individual pathologist [Paradiso et al., 2009], [Hanby, 2005], [Dalton et al., 2000]. Additionally, grading is a time consuming and tedious task. Furthermore, the intra-observer reproducibility (capacity to reproduce the same results over the same histopathological images by the same pathologist at different moments in time) and the inter-observer reliability (the capacity to obtain similar, consistent results over the same histopathological images by different pathologists at the same time or different moments in time) inconsistencies are also problems in the breast cancer grading domain [Paradiso et al., 2009], [Steichen et al., 2006] [Paradiso et al., 2005].

The pathologists now generally use the Nottingham Grading standard system (combining tubule formation, nuclear pleomorphism and mitosis count criteria when analyzing microscopic histopathology images) which provides some guidelines, but taken alone, it does not stand the subjectivity in decision making. These are obviously problems that we propose to address.

A solution to tackle these problems is the Content-Based Image Retrieval approach [Datta et al., 2008], [Long et al., 2003]. The advantages it has for enhancing the intrinsic functionalities of indexing the image features thus to provide reliable image analysis results favored the development of CBIR medical systems. Such examples are PACS (Picture Achieving and Communication System) with the extended version cbPACS [Traina et al., 2005], IRMA [Lehman et al., 2006] and MedGift [Hidki et al., 2007]. In BCG, automated image indexing techniques in the context of CBIR (or independent of) have been developed considering only individual criteria. Nuclear pleomorphism detection and scoring was proposed by [Demir and Yener, 2005], [Jeong et al., 2005], [Adawi et al., 2006], while tubule formation score was addressed by [Petushi et al., 2006] and mitosis count by [Beliën et al., 1997]. Yet, no attempt has been done to combine all criteria in order to provide a complete automated BCG.

Furthermore, the promising development on CBIR in the scientific community did not accrue in the same manner in clinical applications. The reason for such a lack is attributed firstly to the complexity of medical application and secondly to various

gaps of CBIR comprehensively described by [Smeulders et al., 2000], [Müller, 2004], [Deserno et al., 2007], [Datta et al., 2008]. One such *gap is the semantic gap*, defined as the chasm between low-level features of images and high-level concepts semantics. This implies that it is necessary to have meaning associated with the features of objects from the medical images. It is also entangled with the issue of quantitative and qualitative representation, since CBIR indexes image features in a quantitative fashion rather than qualitative.

Crossing roads with Artificial Intelligence field, Case-Based Reasoning (CBR) is another candidate proposed to overcome the problems of medical applications realm. Whilst it offers advantages over CBIR, such as resemblance with the medical reasoning, knowledge representation and management, duality of subjective and objective knowledge or cognitive adequateness, it also has some problems in adaptation from one application to another (similar to CBIR) or concentration on reference [Nillson and Sollenborn, 2004], [Schmidt and Gierl, 2001], [Holt et al., 2006]. Based on these considerations, CBR approach alone is not suitable for BCG.

A logical question arises then: what might be the appropriate solution to formally represent and reason in BCG? Could be CBIR, CBR, or a combination of CBIR with CBR such that to benefit of the advantages of each of them? Decision trees are explored and proposed in a CBR system for retrieving patient files against a case query [Quellec et al., 2008]. Another approach assessed to diabetic retinopathy follow-up and screening mammography and based on a committee of decision trees capable to retrieve possible incomplete medical cases (which are composed of images and semantic content), is proposed in [Quellec et al., 2010a]. Additionally, [Quellec et al., 2010b] advances another CBIR system based on an optimized wavelet transform such that it can be tuned to any pathology and image modality. Or should we take a similar approach to the one of [Jurisica et al., 2001] in which CBIR is integrated into a CBR framework for the protein crystallization application in molecular biology?

To answer this question, another direction is studied: the ontologies, which are currently the most advocated way of representing the knowledge from domains of real world [Chandrasekaran et al., 1999]. In point of fact, this is connected with the advent of semantic web in the universe of World Wide Web, whose main purpose is to capture *semantics* in *a structured formalized way*, understandable by both humans and machines [Shadbolt et al., 2006].

In medical applications, ontologies developed rapidly due to the resemblance with the medical procedure the physicians use. They develop their own languages to help them store and communicate general medical knowledge and patient-related information efficiently.
Different approaches have been taken from non-logic based semantic networks in UMLS, SNOMED-CT [Bodenreider and Zhang, 2006], [Ouagne et al., 2005] to logic-based formalism using OWL [Bontas et al., 2004], and OWL -DL [Golbeck et al., 2003], [Wang and Parsia, 2008], [Zhang et al., 2006]. The advantages that the logic-based formalism offers are the high expressivity of the representation and the reasoning powers in the same time (to ability to obtain a decidable, consistent representation).

Spatial representation and reasoning in biomedical ontologies is also important, as it deals with spatial relations among the concepts from the medical image interpretation and prognosis. Traditional qualitative theories such as mereology (the first-order theory that is concerned with parthood relations), topology (theory concerned with location relations) or mereo-topology [Cohn and Renz, 2008] orientation and distance [Brageul and Guesgen, 2007], [Burrieza et al., 2009] have been proposed for biomedical ontologies as well. [Donnelli et al., 2005] for instance, gives a comprehensive analysis of mereology and location relation in FMA and GALEN, emphasizing the need for *a formal theory support*. One rationale for advocating a formal theory support is to overcome the problem of *ambiguities and inconsistencies in representation*, as discussed by [Schulz et al., 2005] and [Mechouche et al., 2009].
In [Hudelot et al., 2006], the focus is set on topological and metrical relations, whereas [Mezaris et al., 2004] handles geometrical descriptors.

In order to achieve even higher expressivity, [Golbreich et al., 2005] goes further by *proposing ontologies with SWRL rules* for a brain anatomy ontology and investigates the reasoning support that is required. Additionally, [Mechouche et al., 2009] proposes a semantic annotation of gyri and sulci parts of the brain MRI images using OWL-DL and SWRL rules on mereo-topological and orientation relations.

However the problem with rules is that although *they offer expressivity power, there is a trade-off with computational power*; an ontology might end up not being decidable (the reasoner is not able to terminate in a finite time and find a consistent model). To grapple with this issue *SWRL DL-safe rules* have been proposed in the literature [Parsia et al., 2005], [Boley et al., 2005].

Going back to our main thread, we can now give the answer to the question on CBIR-CBR in relation with ontologies. Ontologies represent a reliable approach to narrow the semantic gap and the sharing and reusability characteristics overcome the adaptation problem from the CBR, or the context gap from CBIR. With ontologies it is also possible to formally represent all criterion, from the standard system used in BCG, not only individual criterion as it is the case with image processing algorithms developed so far (which we mentioned above). With the help of ontologies, the subjectivity among physicians can be alleviated. As one can see, ontologies provide even more advantages: high expressivity, formal reasoning, a spatial representation of object features, semantic rules instead of low-level feature-oriented rules. In point of fact, ontologies can connect with CBIR. [Wang et al., 2006] proved that ontologies do help and improve the image retrieval process by narrowing the semantic gap. In the same vein, another work proposed an ontology-based image annotation and retrieval approach [Styrman, 2005], while [Vacura et al., 2008] provides an alternative to MPEG7 standard, by developing a COMM ontology to describe low-level features of images. A histological image retrieval based on semantic content analysis was proposed in [Tang et al., 2003].

Thus in light of the arguments from above, CBIR-CBR combination to provide a new framework is an innovative conceptualization for this BCG medical domain, not discussed elsewhere for BCG, except here in this thesis. Like we mentioned previously, there are some works on CBIR combined with CBR characteristics on molecular biology [Jurisica et al., 2001], on diabetic retinopathy follow-up and

screening mammography [Quellec at al., 2008], [Quellec et al., 2010a], but to our knowledge there is none for breast cancer grading, in particular. It should be mentioned that the methods proposed by [Quellec et al., 2008], [Quellec et al., 2010a-b] could be applied to any pathology and image modality, hence to histopathology as well. However, our novelty implies ontology. Therefore, we propose an ontological approach for the BCG representation that stands at the core of this CBIR-CBR combination in a cognitive virtual microscope system as bona fide prognosis assistance of breast carcinoma.

1.2. Thesis Objectives

In this dissertation we aim at proposing a novel formal representation and reasoning approach for breast cancer grading. In order to reach the goal, our main directions of research and objectives are:

- analysis of different approaches on content and semantic indexing and retrieval in medical applications, CBIR and CBR namely.
- a novel breast cancer grading ontology (BCGO) model using OWL-DL and SWRL formalisms that narrow the semantic gap, offer high expressivity and maintain decidability power. To our knowledge there is no other attempt either in the medical community or in the scientific domain to represent this real world knowledge.
- a spatial theory representation for spatial relations of breast cancer grading domain combining mereo-topology, metric and geometric relations.
- a qualitative evaluation method for the ontological representation of BCGO complemented with medical validation.
- an innovative perspective on virtual microscopy. This approach consists of propelling virtual microscopy to cognitive virtual microscopy based on the integration and guidance of ontology in the image semantic annotation, exploration and retrieval. Furthermore, a cognitive virtual microscope following a CBIR-CBR fusion methodology is another innovative objective we aim at achieving.

1.3. Thesis Structure

The thesis is structured in 9 chapters, first three chapters setting the theoretical foundation for the next three practical chapters focusing on modeling the BCG formal representation. Chapter 7 illustrates model applicability in the context of cognitive virtual microscopy followed by evaluation and validation of the proposed approach. Thesis concludes in chapter 9 with its main contributions and future work.

Chapter 2 gives an introduction on knowledge representation and reasoning, one of the fundamental concepts in Artificial Intelligence field. Based on the definition given in [Davis et al., 1993], we carry out a study on image representation and semantic representation. In the image representation section, we shade light on two different approaches yet similar on their demarches: CBIR and CBR. In semantic representation section, we introduce three formal languages DL, OWL and SWRL, highly promoted in the semantic web. We further show our interest on a particular type of representation that deals with both images and concepts- the spatial representation.
All these studies on CBIR, CBR and on the formal languages are performed with the purpose of identifying a formal representation and reasoning approach for medical applications which are further discussed in chapter 3.

Chapter 3 gives an overview on different approaches of CBIR and CBR in medical applications along with their advantages and shortcomings. We identify one particular problem- the content gap categorized as semantic gap and context gap. We further show a shifting paradigm – from CBIR and CBR to ontologies, which among other characteristics are able to grapple with the content gap. A final analysis on spatial representation in biomedical ontologies is given before closing the chapter.

Chapter 4 introduces Breast Cancer Grading, the microscopic-image based prognosis application we are interested in and the rationale for this interest. We emphasize the need for a formal representation of this domain knowledge and in light of the discussions from chapter 2 and 3, we propose an application-related ontological approach, the Breast Cancer Grading Ontology (BCGO) based on a generalized methodology. The breast cancer grading ontology BCGO aims at formalizing perdurants which is highly related with the spatial extension we add to the model.

Chapter 5 focuses on the spatial representation. It presents the formal spatial theory for the spatial relations of breast cancer grading domain. We show that such a support helps in reducing ambiguities and inconsistencies in representation by two means: manual reasoning (which resembles with the medical reasoning) and formal reasoning based on Description Logics tableau algorithm.

Chapter 6 proceeds with a detailed presentation of the breast cancer grading ontology implementation in the Protégé framework based on the methodology proposed in chapter 4. We discuss specific characteristics of OWL such as Open World Assumption, defined and primitive classes. DL related aspects are described in connection with SWRL, in order to achieve high expressivity and satisfiability of the BCGO model.

Chapter 7 treats the aspects of evaluation and validation of the BCGO model. We propose a qualitative evaluation approach using qualitative metrics complemented with DL evaluation in terms of syntactic constraints. We emphasize the importance of the expressivity of the formal language for the reasoning tasks. Semantic retrieval method and medical feedback provide the means for the validation of the ontology. The integration of our ontology in the biomedical portal stresses the reusability purpose of ontology's development.

Chapter 8 introduces a model applicability context from a cognitive virtual microscopy perspective. We develop a cognitive microscope prototype MICO in which the formal semantic representation encapsulated in the BCGO plays a key role as it provides semantic annotation and retrieval support in the grading. This represents a significant asset in terms of cognitive vision. We also show that MICO follows the combined CBIR-CBR methodology proposed in chapter 2.

Chapter 9 summarizes the contributions brought in this thesis followed by future research direction in the formal representation of breast cancer grading domain.

2. Knowledge Representation and Reasoning

Knowledge representation (KR) is considered one of the fundamental concepts in Artificial Intelligence. Whilst there has been a lot of argue over the years on what ways of representation should we adopt, or what features a representation should have, many fail to address the problem of what representation exactly is. In [Davis et al., 1993], the authors tackle this issue by giving a fundamental mindset of five key roles a knowledge representation is defined of. In synthesis, these five characteristics are:

- A knowledge representation is essentially *a substitute for the thing itself*, by which an entity thinks instead of acting, in order to determines consequences (1)
- It is *a set of ontological commitments*, i.e., an answer to the question: In what terms should I think about the world? (2)
- It is a *fragmentary theory of intelligent reasoning* conveyed by: (i) the fundamental conception of intelligent reasoning; (ii) the set of inferences the representation supports; and (iii) the set of inferences it recommends (3)
- It is a *medium for pragmatically efficient computation*, i.e., the computational environment in which thinking is performed (4)
- *It is a language in which we say things about the world* (medium of human expression -5)

They also note that the implications of such a knowledge representation view are manifold. It firstly gives flexibility on what type of representation one should choose as long as it follows these guidelines. In other words, we can encompass a wide variety of representations. It follows logically that a representation can have different properties depending on which role it fulfills. Secondly, representations can be combined in a way that they do not contradict the definition. Lastly, the representation and the reasoning are intertwined. An explicit representation of a knowledge corpus is connected with its logical consequences obtained through the process of inference. In this setting, that is the meaning of intelligent reasoning [Davis et al., 1993].

Given this basis, we argue that *an image representation also falls into the category of knowledge representation* even though it is not explicitly part of artificial intelligence domain. Image representation is a language in which we express things about the world in which we perform computations and it is an image-based reasoning. Image representation is also a very important aspect of the world. It thus fulfills roles 3 to 5.
To this end, one of the representations we refer to is the Content-Based Image Retrieval (CBIR). This view of CBIR from the knowledge representation perspective is also taken by [Sciacio et al., 2002].

Next sections give our theory on why CBIR is a knowledge representation. We further study another type of representation: the Case-Based Reasoning and we advance an approach to connect them.
On this foundation, we propose a method to link images with semantics. It is important to mention that from the perspective of the roles of knowledge representation, neither CBIR, not CBR have to meet all five roles in order to be qualified as knowledge representation, since there is a wide variety of representation.

2.1. Image Representation

We introduce generic considerations on the concepts of CBIR and CBR in terms of methodology/technology context, issues and techniques used [Tutac et al., 2009a]. Identifying similarities and differences between the two approaches at the indexing, retrieval and relevance feedback/case adaptation level with reasons for this kind of attempt is the next objective of our discourse.

2.1.1. Content-Based Image Retrieval

Content- Based Image Retrieval (CBIR) is generally seen as a technology using content-similarity-based methods to solve problems, particularly, to access pictures from image database by visual content according to the users' interest [Long et al., 2003], [Datta et al., 2008]. Hence, *it is a language in which we say things about the world, in particular related to the images of various things of the world.* Therefore, CBIR fulfills role 5 of a knowledge representation.

In the first approaches, CBIR consists of two main phases: the indexing and the retrieval typically based on visual similarity. More recently, the relevance feedback has been perceived as an integrated part of the CBIR demarche, yet considered as a key-issue in CBIR [Smeulders et al., 2000]. From its early stages of theoretical foundations and first systems development attempts hitherto, CBIR has opened promising perspectives in the research area. The reasons are manifold, but foremost the idea of achieving valid retrieval results when given a query challenged research community to define advanced indexing and retrieval techniques. It thus enriches the core functionality of organizing increasing digitized data. By which means solving problems with respect to types of query, similarity computation, relevance of results retrieved and so forth, are still opened questions CBIR faces in its development, moreover on medical field.

How does a CBIR work? Using various query mechanisms, such as Query-By-Visual-Example (QBVE) the user presents a sample image, an image region of interest or a pattern of the image, to the system that tries to respond with similar images related to the given query [Zhao and Groski, 2001]. Each image from the database is indexed in the indexing phase (offline- the purple box), according to its signature extracted as a discriminant numerical features vector. In the retrieval phase (online- the beige box), the query signature is compared to the repository signatures and similar retrieved images are further used in a relevance feedback step to refine the query in order to improve system performance (recall metric).

It can be said that CBIR is a *medium for pragmatically efficient computation* (role 4), as the indexing and retrieval of features implies computation of signatures and computation of similarities between signatures of images.

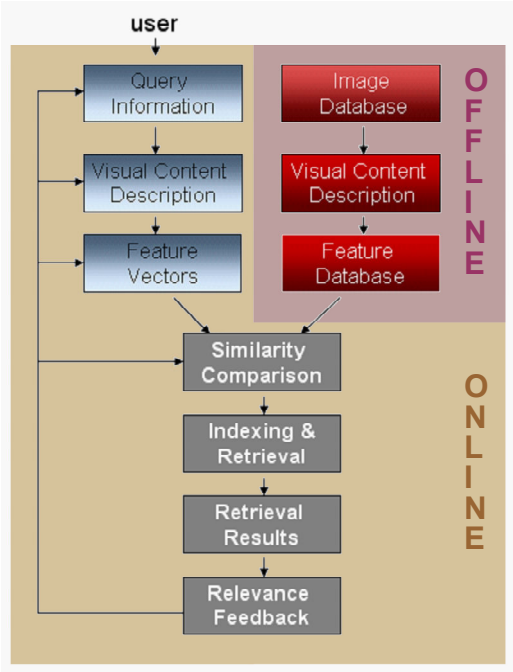


Figure 2.1. CBIR functional diagram

The key idea of CBIR is the visual content processing and analysis. The mechanism of inference is based on the visual content of the image. The set of inferences the representation supports are viewed in terms of image salient features which are indexed. The relevance feedback phase contains the set of inferences recommended in order to refine the query. It can thus be called *an image-based reasoning* approach (role 3 of a knowledge representation).
The new trends of using high level semantics concepts combined with the low level visual features for an efficient indexing and retrieval have been discussed in the literature [Little and Hunter, 2004], [Kalfoglou et al., 2006], [Liu et al., 2004], [Carneiro et al., 2007] and the issues of CBIR regarding this attempt will be further detailed, with predilection to the medical applications.
We consider that such an approach has deep implications if used in medical applications, being able to provide more effective diagnosis and prognosis assistance. To illustrate the CBIR demarche, a functional diagram is presented in Figure 2.1, adapted from [Long et al., 2003], which also illustrates role 4 and 5 of a knowledge representation.

Related to the organizing of digital visual data, one of the main characteristics of CBIR is the single-image-way of structuring information. This characteristic becomes an issue when developing medical CBIR due to the different way patient information is usually structured: by cases.

Another issue is related to the necessity to start from scratch the whole relevance feedback process at each new query, without having the possibility to capitalize the past experiences and the related knowledge.

Thus, we introduce CBR as a solution to this problem based on its characteristic to structure information by cases.

2.1.2. Case-Based Reasoning

Also designed for problem solving but from another perspective, Case-Based Reasoning has been proposed as an efficient approach in Artificial Intelligence, particularly in Knowledge-Based Systems [Watson and Marir, 1994] (KBS- see Figure 2.2). Similarly to CBIR, CBR also represents *a language to describe things of the world,* in various applications (role 5 of a knowledge representation).

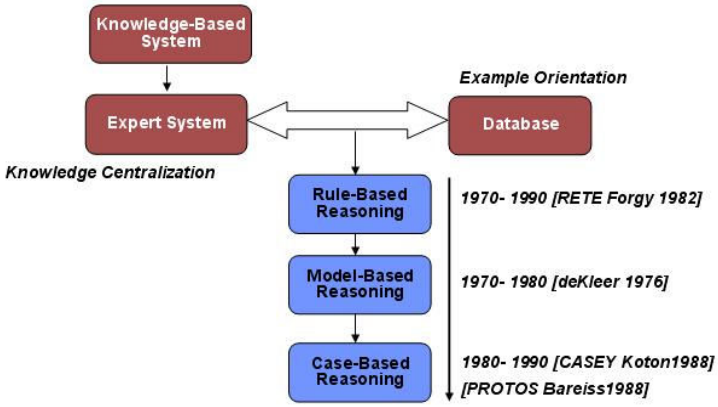


Figure 2.2.CBR systems origin

An excellent review of CBR foundations, techniques and system/tools developed is given by [Aamodt and Plaza, 1994], [Watson and Marir, 1994], and [Pal and Shiu, 2004]. Although different propositions have been made for KBS, many issues remained unsolved. CBR arose as a promising solution to tackle with KBS specific drawbacks, as illustrated in Table 2.1.

KBR paradigms [Adawi et al., 2006], [Watson and Marir, 1994], [Richter, 2003]	**Advantages**	**Drawbacks**
Rule-Based Reasoning (RBR)	➢ appropriate when domain knowledge is fully available ➢ simple to design & implement	➢ time consuming knowledge-acquisition task ➢ formalization/abstraction of knowledge ➢ insufficiency of cases to extract a domain model
Model-Based Reasoning (MBR)	➢ efficient when a generic model is described ➢ appropriate for small volume of information	➢ availability of a model ➢ time & management ➢ implementation &maintainability ➢ knowledge elicitation bottleneck
Case-Based Reasoning (CBR)	➢ elicitation doesn't require any explicit model ➢ database management techniques ➢ knowledge is inserted by learning new cases ➢ signature features for a case description ➢ experience-based approach ➢ incrementally development	➢ possible inexactness when the actual solution is not identical with the previous one

Table 2.1. KBR paradigms

From the KBR paradigms, CBR distinguishes by its elicitation characteristic and the way knowledge is handled. Firstly, in a CBR system there is no need of an explicit model, unlike RBR and MBR. Hence, it is a proper solution for complex applications. Secondly, it is not required to have complete domain knowledge when a CBR system is constructed. The knowledge is inserted by learning each new case in an incremental fashion. This helps the development of a sophisticated CBR system, starting from minimum knowledge required.

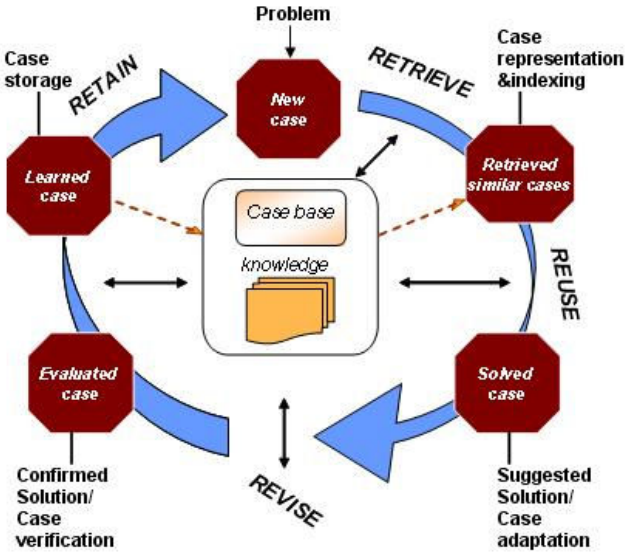


Figure 2.3. CBR REs cycle

The hallmark of CBR is working with structured information by cases and trying to retrieve the similar cases based on past experience when given a new problem. Guided by the Pareto principle, the intrinsic characteristic of CBR is the reasoning by remembering previous stored experiences [Leake, 1996]. Essentially, it is defined as the four REs cycle: Retrieve, Reuse, Revise and Retain as main principles, each of them having particular phases [Aamodt and Plaza, 1994]. Figure 2.3 gives a clear view of how CBR functions.

In order to do the retrieval, case representation and case indexing phases are required. The case representation and indexing phase implies computation of the information from the cases such that is possible to retrieve cases, by similarity computation. Therefore, role 4 of a knowledge representation is fulfilled.

The retrieved cases are then reused to provide a possible solution to the given problem and therefore, a case adaptation phase is considered. In the revise step, the proposed solution is evaluated in terms of its applicability in the real-world. The solution is finally retained as a part of the new case and thus, the case-base is updated with each new learned case for future problem solving. These phases confirm role 3 of this representation, as they function based on a set of inferences or rules that need to be applied to evaluate the solution and to retain it. The set of inferences or rules are defined based on the type of application. Figure 2.3 illustrates as well role 3 to 5 of CBR as a knowledge representation.

The issues of CBR and their possible solutions will be explored in chapter 3, oriented on the medical axis.

2.1.3. Methodology or Technology?

In order to define our analysis strategy, we firstly make a distinction between methodology and technology applied to CBIR and CBR. Technology and methodology can have different meanings in different contexts. Webster's dictionary defines technology as:

***Definition 2.1. Technology** is:*

1. *the practical application of knowledge especially in a particular area (i.e. medical technology), or the capability given by the practical application of knowledge*
2. ***a manner of accomplishing a task especially using technical processes and methods or knowledge***
3. *the specialized aspects of a particular field of endeavor*

Webster's dictionary gives the definition of the methodology as:

***Definition 2.2 Methodology** is*

1. ***a body of methods, rules and postulates employed by a discipline; a particular procedure or a set of procedures***
2. *the analysis of the principles or procedures of inquiry in a particular field*

Conceptually, CBIR is often referred as a technology that uses various techniques to solve specific problems [Long et al., 2003], [Datta et al., 2008] and [Smeulders et al., 2000] (definition 2).
Similarly, [Kolodner, 1993], [Richter, 2003] consider CBR as a technology, whilst [Watson, 1999] stresses that CBR is an organized *set of principles* which guide action in problem solving matters rather than an isolated technique, limited to handle only very specific tasks. Hence, it verifies the first definition of Webster's dictionary and also confirms the definition of a methodology given by [Checkland and Scholes, 1990].
The reason for viewing CBR as a methodology launches deep implications. On one hand, since it doesn't have its own technology, it can use *any* technology that applies CBR principles. On the other hand, we can build *hybrid systems,* in terms of *hybrid methodologies* and not hybrid technologies.
Furthermore, seeing CBR as a methodology supports the idea of future research, which is important, since many technologies for each CBR phase are commonly used and some -already mature. We adopt the same approach of Watson's and moreover, we propel CBIR at the same level, of methodology. We envisage that CBIR made significant development in the recent years in terms of concepts, techniques and application domain. In our opinion, CBIR of today does not only organize digital data, as it was the main objective in its early years due to the semantic web development. In this setting, we consider that the principles of CBR can also define CBIR as a methodology, except the last principle (Retain). It is not necessary to have all four principles of CBR to see CBIR a methodology; the key idea is to define CBIR in terms of basic principles and use any techniques in line with its principles. Our general paradigm is depicted by Figure 2.4. Methodology is defined by principles that are further used when applying various technologies to the given application.

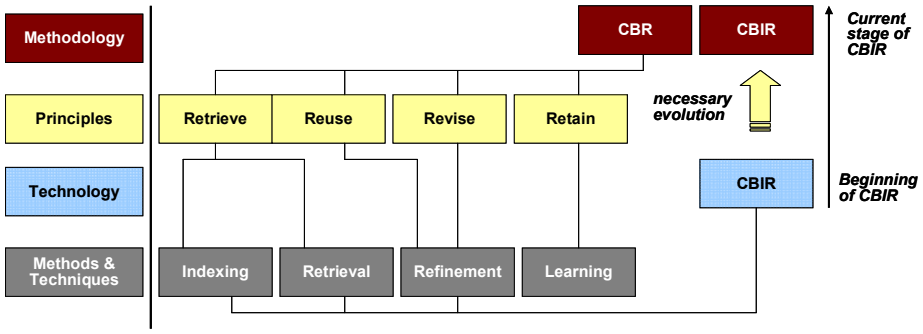


Figure 2.4. CBIR & CBR. Methodology versus Technology

Each technology comes with its set of combined methods and techniques. The reasons and the implications for CBIR and CBR technology versus methodology are detailed in Table 2.2.

Methodology versus Technology	Definition	Reason	Implication
Content- Based Image Retrieval	Technology [Long et al., 2003], [Datta et al., 2008]	*set of methods* to solve problems [Smeulders et al., 2000]	➢ manifold applications ➢ multitude of gaps issue
	Methodology	continuous development	➢ knowledge-based systems flexibility
Case-Based Reasoning	Technology [Kamp et al., 1998], [Richter, 2003]	Artificial Intelligence technology description [Kamp et al., 1998]	➢ task limitation ➢ research limitation
	Methodology [Watson, 1999]	*set of principles* to solve problems [Watson, 1999]	➢ can use any technology ➢ hybrid systems ➢ future research

Table 2.2. Reasons and Implications for Methodology/Technology in CBIR & CBR

To illustrate how CBIR follows the principles of CBR methodology, we construct the Res cycle based on the CBIR functional diagram (Figure 2.5).

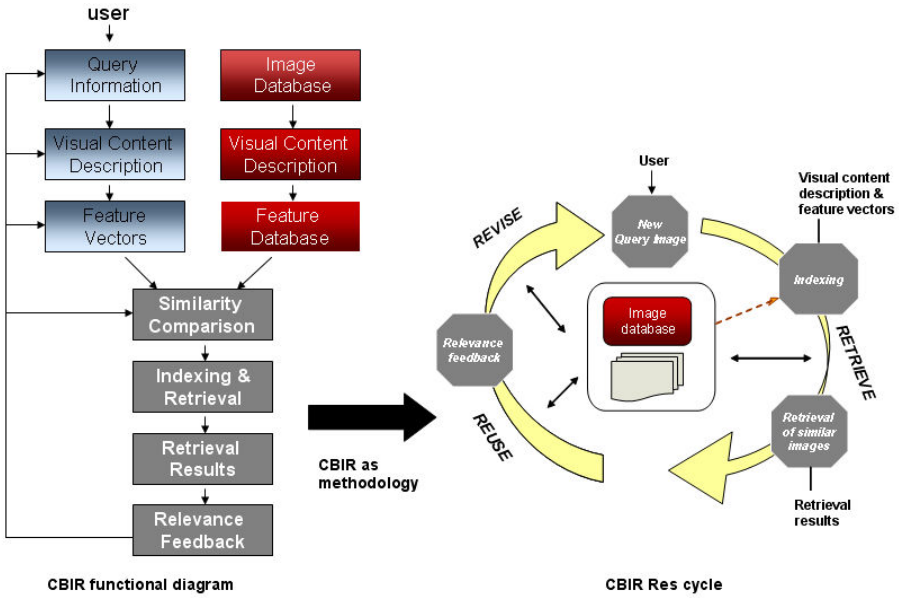


Figure 2.5. CBIR methodology

2.1.4. Comparative Analysis of CBIR and CBR

Since we consider both CBIR and CBR as methodologies, the aim of this chapter is to present the similarities and the differences of three major technologies common for both CBIR and CBR: the indexing, the retrieval and the search refinement. Thus, to benefit of their merits and to overcome their week points for our hybrid framework proposal.
We want to note that we do not go into every detail for each axis and methods employed, as to use the microscopic lens. We rather adopt a generic method of comparison and analysis to fit the purpose of this thesis. In this sense, our discussion on CBIR and CBR could be seen as limited. However, this approach can be further used as a support for other works.

Indexing in CBIR & CBR

As shown in Table 2.3, the core distinction is the principle-based orientation of CBR, unlike CBIR that was generally considered technique-based oriented until now.
At the techniques level, there are some similar works proposed (learning-based in CBIR and inductive-learning in CBR) as well as some different approaches (for instance, no correspondence in CBIR for similarity & explanation-based technique found in CBR).

Indexing	**CBIR** [Smeulders et al., 2000], [Long et al., 2003], [Datta et al., 2008]	**CBR** [Watson and Marir, 1994], [Adawi et al., 2006], [Richter, 2003]
Indexing Principles	—	➢ predictive ➢ purpose oriented ➢ abstract/concrete enough
Similar Indexing Techniques	➢ feature-based ➢ structural features	➢ features &dimensions
	➢ salient-features	➢ difference-based
	➢ learning-based	➢ inductive learning
Different Indexing Techniques	—	➢ similarity & explanation-based
Characteristics	➢ feature indexing	➢ case indexing

Table 2.3. Indexing in CBIR & CBR

Semantic indexing represents another category related to our analysis. At this point, an important issue identified is the semantic gap, defined as the discrepancy between the low level visual features and the high-level semantic concepts [Smeulders et al., 2000].
The essence is the fact that this gap is biased by the versatility of visual image content. Most common research into bridging the semantic gap is actually tackling the descriptors and object labels level and some approaches focus on extracting the objects from a raw image to label them [Smeulders et al., 2000], [Long et al., 2003] and [Müller, 2004]. Hence, the cue of semantics is to find the means and knowledge to associate meaning to some features retrieved from the image (to index images by semantic means) according to a specialized or generalized knowledge [Vasconcelos, 2007].
The new trend is to design and model ontologies that can work at the semantic level with domain knowledge support, thus emphasizing the solution of apriori knowledge injection. Hence, the similarity with CBR design principle can be identified here. Furthermore, we advance the idea that ontologies, proposed as an interesting perspective to narrow the semantic chasm in the scientific literature [Smith, 2004], can also be applied to CBR, in the sense of structuring and representing knowledge contained in the cases. At this point, there is an elaboration to be done. Ontologies are defined as a formal and explicit specification of an abstraction [Gruber, 1993]. Similarly, CBR knowledge is explicitly stored in concrete cases, thus implying the fact that the case is not a general rule, but an instantiation of a formal specification.

Retrieval in CBIR & CBR

A similar analysis is applied to retrieval phase in CBIR and CBR. Once again, the need to have guiding principles in CBIR is pointed out, in Table 2.4. Modeling similarity is a central element to navigate through the space of possible solutions, in

CBIR as well as in CBR, different approaches depending on image indexing/case representation.
A retrieval based on semantic example in CBIR is correspondently found in CBR as a knowledge-guided retrieval. One of the most common retrieval techniques successfully applied in both fields is the Nearest-Neighbor Retrieval. From the dissimilarity point of view, there are some techniques that are used either in one or in the other (for instance, validated retrieval in CBR, or Query-by-Keyword in CBIR). To this end, a classification of CBR into two categories is to be considered. Most CBR systems fall in the problem- solving category, which uses previous cases to only suggest the most likely solution to be applied to the new case. In contrast, interpretive CBR are based on reference cases – previous cases, per se, to solve the new problem [Pal and Shiu, 2004]. In the same paper, a summary of soft CBR, implying combination with AI techniques is given to emphasize the idea of methods integration when it comes to evaluate the results from the reliability standpoint.

Retrieval	**CBIR** [Müller, 2004], [Vasconcelos, 2007]	**CBR** [Watson and Marir, 1994], [Adawi et al., 2006], [Richter, 2003]
Retrieval Principles	___	➢ criteria selection ➢ memory model
Similar Retrieval Techniques	➢ Query-By-Semantic-Example ➢ semantic retrieval	➢ knowledge-guided
	➢ Nearest-neighbor retrieval	➢ Nearest-neighbor retrieval [Kolodner, 1993]
Different Retrieval Techniques	➢ Query-by-Keyword ➢ Query-by-Visual-Example(QBVE) [Little and Hunter, 2004]	➢ inductive ➢ validated retrieval

Table 2.4. Retrieval in CBIR & CBR

In essence, the process of retrieval highly depends on the indexing phase and the similarity computation step. The higher the efficiency of indexing, the better the retrieval. The same illation is also found with respect to medical applications. Table 2.4 gives an overview of retrieval in CBIR and CBR.

Refinement in CBIR & CBR

Relevance feedback is defined as supervised active learning query modification/adaptation technique to improve the effectiveness of the information systems [Datta et al., 2008]. Likewise, case adaptation of CBR reuse principle focuses on refinement of the proposed solution of the similar cases extracted at retrieval time. Our rationale to consider relevance feedback and case adaptation as correspondent is due to their basic idea: refinement. The difference between the two approaches appears with respect to the *target of refinement*: in CBIR, *the query* is to be refined (to improve the response), while in CBR, *the solution* is refined.

The performance of the system is evaluated using two metrics, known as precision and recall, with values between 0 and 1.

$$Recall = \frac{Number\_of\_relevant\_images\_retrieved}{Total\_number\_of\_relevant\_images\_in\_the\_database} \quad (2.1)$$

$$Precision = \frac{Number\_of\_relevant\_images\_retrieved}{Total\_number\_of\_images\_retrieved} \quad (2.2)$$

Another alternative to formulate precision and recall is in terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). True positive corresponds to retrieved images that are indeed relevant, true negative means images that are not retrieved and are not relevant, false positive- images that were retrieved and should not have been relevant, false negative- images that were not retrieved but should have been relevant.

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

Relevance feedback is nowadays regarded as a vital need and captured a high interest in the CBIR development. A relevance feedback step integrated in the CBIR process has several implications. Firstly, it is possible to create the link between the low level features and the high level concepts, capturing user and query specific semantics. Secondly, relevance feedback improves both precision and recall, but when it refines the ranks accordingly to the query adaptation, it improves the system recall [Manning et al., 2008].
There are however, some drawbacks such as increasing of the user involvement (multiple rounds of feedback affect user's patience) or the fact that the changes would not be done at the low level features (they will remain the same). Additionally, human perception of image similarity is highly subjective, task-dependent and it is sometimes hard to establish why the obtained images are similar and how to exactly improve the performance of the system. Hence, it is necessary to have a relevance feedback in a CBIR system. Yet, some approaches provide no relevance feedback or a naïve feedback. An alternative solution would be an offline training as illustrated in Figure 2.1 as well, which will also solve the user involvement issue.
Similar situation can be encountered in a CBR system, the so-called null adaptation technique. The analogy between the techniques used in CBIR and in CBR, as well as the single-oriented methods is described in Table 2.5.
To conclude this section, a strong point for the CBR regards its closed loop characteristic. CBR process does not stop at the retrieval phase as it is most likely to happen in a CBIR system; it goes further to the adaptation of the solution and moreover the new case, after revision (if necessary) is stored in the case-base. Thus the CBR is incrementally learning, the knowledge is continuous expanding, unlike CBIR where there is no such step beyond.

Relevance feedback RF/Case adaptation [Long et al., 2003]	**CBIR** [Long et al., 2003], [Zhao and Groski, 2001], [Datta et al., 2008]	**CBR** [Watson and Marir, 1994], [Aamodt and Plaza, 1994], [Richter, 2003]
RF Principles	—	➢ structural [Kolodner, 1993] ➢ derivational
Similar RF/Case Adaptation Techniques	➢ no RF/naïve RF	➢ null adaptation
	➢ feature re-weighting	➢ parameter adjustment
	➢ specialized user-driven	➢ critic-based
	➢ memory-retrieval	➢ model-based
	➢ active-learning	➢ abstraction & respecialization/reinstantiation
Different RF/Case Adaptation Techniques	➢ probabilistic [Vasconcelos, 2007]	➢ derivational replay ➢ case base substitution
Characteristics	➢ query refinement	➢ solution refinement

Table 2.5. Relevance feedback/Case adaptation in CBIR & CBR

Therefore, we consider CBIR as an open loop; even if there is a week relevance feedback, the process starts over again and thus, there is no recording of how the problem was solved in the past.

CBIR	***CBR***
➢ image based	➢ textual information based
➢ limited to retrieval phase	➢ integrated new case after adaptation & revise
➢ query expansion	➢ no query
➢ lack of knowledge injection	➢ a priori knowledge
➢ semantic web-based	➢ knowledge-based
➢ structured by individual element	➢ structured by cases
➢ weak learning/static database	➢ incrementally learning/dynamic database
➢ context-dependent	➢ context-modeling
➢ open loop : naïve relevance feedback	➢ close loop: case adaptation-case storage

Table 2.6. CBIR and CBR

However, there are also some tradeoffs at the case storage level of CBR. Storing too many cases may affect the speed of the execution and may introduce overfitting problems. To face this issue, a rough set theory (RST) combined with formal concept analysis (FCA) was proposed in [Tadrat et al., 2007].
Table 2.6 contains the outline ideas of the main differences between CBIR and CBR.

2.2. Semantic Representation

In this section the spotlight is set on the link between the language of images and the formal languages particularly used for representation and reasoning on the concepts of the application domain. We address the Description Logics formalism which is further connected with ontologies and their language. In the last part we introduce a language for rules which is important in the demarche of the thesis.

2.2.1. Description Logics Formalism

The family of Description Logics represents the knowledge of an application domain using a set-theoretical foundation [Baader et al., 2003a], [Baader et al., 2007]. The Description Logics models concepts, role and individuals with the help of operators such as existential and universal operators to make the language decidable and of low complexity.
Concept names are equivalent to unary predicates, role names to binary predicates and individuals are associated with constants.
Like any language, it has syntax - the collection of symbols which are legal expressions within the domain (Δ) - and semantics ($\mathcal{I}$), which determine the meaning of the symbols.
We will present the characteristics of the basic DL language, the $\mathcal{ALC}$ (Attributive Logics with Complements). The name of this DL is given by the complement of the atomic concept C ($\neg C$) or the negation of any concept allowed, not only atomic concept.

DL syntax and semantics

According to Tarsky style [Stroińska and Hitchcock, 2002], the domain and its interpretation function are noted as

$$(\Delta^{\mathcal{I}}, .^{\mathcal{I}}) \tag{2.5}$$

$\Delta^{\mathcal{I}}$ is the domain, a non-empty set

$.^{\mathcal{I}}$ is the interpretation function that maps:

concept name (class name) *C* -> subset $C^{\mathcal{I}} of \Delta^{\mathcal{I}}$

role name (property name) *R* -> binary relation $R^{\mathcal{I}} over\ \Delta^{\mathcal{I}}$

individual name *i* -> $i^{\mathcal{I}} element\ of\ \Delta^{\mathcal{I}}$

With these concepts, roles and operators which represent the grammar of $\mathcal{ALC}$ language, one could construct complex concept *expressions*. From concepts expressions to terminological axioms, two other operators are introduced:

subsumption and equivalence, noted as in the following formulas. The interpretation (the meaning) $\mathcal{I}$ of concept subsumption is that concept *E* is a necessary condition for concept *C* in order for *C* to be subsumed by *E*. Similarly for the interpretation of the concept equivalence, *E* is a necessary condition for concept *C* in order for *C* to be equivalent with *E*. We will also discuss concept subsumption and concept equivalence at the DL reasoning tasks.

$$C \sqsubseteq E \quad \text{(Concept subsumption)} \tag{2.6}$$

$$C \equiv E \quad \text{(Concept equivalence)} \tag{2.7}$$

We illustrate the concept subsumption and concept equivalence by two simple examples: conference is subsumed by event and a woman is equivalent to a female human.

$$Conference \sqsubseteq Event$$

$$Woman \equiv Human \sqcap Female$$

Table 2.7 gives a description of $\mathcal{ALC}$ syntax and semantics with concept, roles, operators and concept expressions.

DL Knowledge base

A knowledge representation makes use of a knowledge-base $\mathcal{K}$, which in DL is formed by a pair ($\mathcal{T}$, $\mathcal{A}$).

$\mathcal{T}$ stands for terminology *TBox* which contains the definition of concepts (*axioms*) along with the statements of constraints (*roles with regard to concepts*). $\mathcal{A}$ stands for *ABox* and contains all the *assertions* about the individuals (specialization of concepts and roles), divided into concept assertions and role assertions. In other words, concepts describe sets of individuals and roles represent relations among individuals.
*TB*ox axioms are of the form:

$$C \sqsubseteq E,\ C \equiv E,\ R \sqsubseteq S,\ R \equiv S,\ R^{+} \sqsubseteq R \tag{2.8}$$

where *C, E* are concepts, *R, S* roles and R^{+} represents the set of transitive roles.
For instance, a role subsumption states that has_daughter is subsumed by has_child. Role equivalence is illustrated between has_price and has_cost, while has_ancestor indicates a transitive property.

$$has\_daughter \sqsubseteq has\_child$$
$$has\_price \equiv has\_cost$$
$$has\_ancestor^{+} \sqsubseteq has\_ancestor$$

*AB*ox assertions are of the form:

$$x : C,\ \langle x,y\rangle : R \tag{2.9}$$

where *C* is concept, *R* role and *x* and *y* denote the individuals. For instance, Ann is a Women, who has a daughter called Mary. Both Ann and Mary are individuals.

$$Ann : Women,\ \langle Ann, Marry\rangle : has\_daughter$$

The DL semantics creates the *model* of structure*.* To give semantics to an *AB*ox, an extension of the interpretation to the individuals is carried out. Hence, the interpretation $\mathcal{I}$ maps atomic concepts and roles to sets and relations, and maps each individual name to an element from the domain of discourse. An *interpretation* is called a *model* if that interpretation satisfies the axioms or assertions from the knowledge-base, which is accomplished during the process of reasoning.

DL Symbols	**Syntax**	**Semantics (Interpretation $\mathcal{I}$)**	**Description**
T	T	$\Delta^{\mathcal{I}}$	top concept (universe)/ all concept names
$\perp$	$\perp$	$\varnothing$	bottom concept (empty set)
	C	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$	atomic concept (unary)
	R	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} x\ \Delta^{\mathcal{I}}$	role concept (binary relation)
	o	$o^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$	individual o
$\neg$	$\neg C$	$(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$	atomic negation (negation operator)
$\prod$	$C \prod E$	$(C \prod E)^{\mathcal{I}} = C^{\mathcal{I}} \cap E^{\mathcal{I}}$	atomic conjunction (intersection operator)
$\amalg$	$C \amalg E$	$(C \amalg E)^{\mathcal{I}} = C^{\mathcal{I}} \cup E^{\mathcal{I}}$	atomic disjunction (union operator)
$\forall$	$\forall R.C$	$\{a \in \Delta^{\mathcal{I}} \mid \forall b.(a,b) \in R^{\mathcal{I}} \rightarrow b \in C^{\mathcal{I}}\}$	the set of all individuals, which iff all take part in relation *R*, are related through *R* to only individuals of concept *C* (value/universal restriction)

$\exists$	$\exists R.C$	$\{a \in \Delta^{\mathcal{I}} \mid \exists b.(a,b) \in R^{\mathcal{I}} \cup b \in C^{\mathcal{I}}\}$	the set of all individuals which are related through R to at least one individual of concepts C
$\sqsubseteq$	$C \sqsubseteq E$	$C^{\mathcal{I}} \subseteq E^{\mathcal{I}}$	concept subsumption
$\equiv$	$C \equiv E$	$C^{\mathcal{I}} = E^{\mathcal{I}}$	concept equivalence
$\models$	$\mathrm{T} \models C \sqsubseteq E$	$\mathrm{T} \models C \sqsubseteq E$ *if and only if* $C^{\mathcal{I}} \subseteq E^{\mathcal{I}}$	T models concept subsumption
$\doteq$	$C \doteq E$		concept definition (concept C is defined to be equal to E)
:	$a : C$ / $(a,b) : R$		concept assertion (individual a is a C) /role assertion (individual a is R-related to individual b)

Table 2.7.Syntax and semantics of $\mathcal{ALC}$ language [Obitko, 2007]

An interpretation $\mathcal{I}$ satisfies a *TBox* axiom $\mathcal{AX}$, $\mathcal{I} \models \mathcal{AX}$ if the following are satisfied (where iff means if and only if):

$$\begin{aligned}
&\mathcal{I} \models C \sqsubseteq E \text{ iff } C^{\mathcal{I}} \subseteq E^{\mathcal{I}} \\
&\mathcal{I} \models C \equiv E \text{ iff } C^{\mathcal{I}} = E^{\mathcal{I}} \\
&\mathcal{I} \models R \sqsubseteq S \text{ iff } R^{\mathcal{I}} \subseteq S^{\mathcal{I}} \\
&\mathcal{I} \models R \equiv S \text{ iff } R^{\mathcal{I}} = S^{\mathcal{I}} \\
&\mathcal{I} \models R^{+} \sqsubseteq R \text{ iff } (R^{\mathcal{I}})^{+} \subseteq R^{\mathcal{I}}
\end{aligned} \tag{2.10}$$

$\mathcal{I}$ satisfies a *TBox* $\mathcal{T}$,$\mathcal{I} \models \mathcal{T}$ iff $\mathcal{I}$ satisfies every axiom $\mathcal{AX}$ from $\mathcal{T}$.

An interpretation $\mathcal{I}$ satisfies an $\mathcal{ABox}$ assertion $\mathcal{AS}$, $\models \mathcal{AS}$

$\mathcal{I}$ satisfies an *AB*ox $\mathcal{A}$, $\mathcal{I} \models \mathcal{A}$ iff $\mathcal{I}$ satisfies every assertion $\mathcal{AS}$ from $\mathcal{A}$.

$$\mathcal{I} \models x : C \text{ iff } x^{\mathcal{I}} \in C^{\mathcal{I}}$$
$$\mathcal{I} \models \langle x, y \rangle : R \text{ iff } x^{\mathcal{I}}, y^{\mathcal{I}} \in R^{\mathcal{I}} \quad (2.11)$$

$\mathcal{I}$ satisfies a KB $\mathcal{K}$, $\mathcal{I} \models \mathcal{K}$ iff $\mathcal{I}$ satisfies both $\mathcal{A}$ and $\mathcal{T}$.

DL reasoning means the ability to infer logical consequences from the explicitly defined set of axioms and assertions, from the explicit knowledge. There are different tasks of reasoning, classified into four main categories: with regard to (noted as wrt) concepts, wrt *TBox*, wrt *ABox* and wrt $\mathcal{K}$, respectively [Horrocks, 2000].

Abstracting TBox is the *TBox reasoning* task that can be fulfilled by proving concept satisfiability. Essentially, it consists of expanding the $\mathcal{T}$ (resulting $\mathcal{T}'$) by expanding concept *C* with *C'* with regard to $\mathcal{T}$.

A straight-forward reasoning task regarding *AB*ox *is the instance checking*, since *AB*ox contains instances of the concepts from classes. An assertion α is entailed by $\mathcal{A}$ iff every interpretation $\mathcal{I}$ that satisfies $\mathcal{A}$ (that is a model of $\mathcal{A}$) also satisfies α.

$$\mathcal{A} \models \alpha \quad (2.12)$$

As *AB*ox contains role assertions and concept assertions, we need to verify instances for both of them. If α is a role assertion, instance checking is not complicated.
If α is a concept assertion *C* (α), instance checking can be reduced to concept consistency.

$$A \models C \text{ iff } A \cup \{\neg C(\alpha)\} \text{ is consistent} \quad (2.13)$$

Knowledge-base satisfiability or *AB*ox satisfiability with regard to *AB*ox implies that $\mathcal{A}$ is consistent wrt $\mathcal{T}$, iff there is a model of both *AB*ox and *TB*ox, which satisfies every axiom from $\mathcal{T}$ and every assertion from $\mathcal{A}$.

$$\mathcal{I} \models KB \quad (2.14)$$

Reasoning tasks about *concepts* deal with four types of verification related to concept inference, consistency of a concept (satisfiability) and concept in relation with other concepts: equivalence and disjointness. It is shown in the specialized literature that all these concepts tasks can be reduced to concept satisfiability. Furthermore, tasks related to *ABox* and *TBox* either alone or defining the knowledge base $\mathcal{K}$, can all be reduced to concept satisfiability. This is logical due to the fact that concept represents the core element of a knowledge-base.

- **Concept subsumption**
  C subsumed by E iff $C^{\mathcal{I}} \subseteq E^{\mathcal{I}}$ *for* $\mathcal{I}$ *of* $T \rightarrow C \sqsubseteq_T E$ *or* $T \models C \sqsubseteq E$

(where $_{\mathcal{T}}D$ denotes concept *D* which belongs to *TBox* $\mathcal{T}$ or that $\mathcal{T}$ satisfies *C* is subsumed by *D*)

- **Concept satisfiability**
  C satisfiable wrt T iff $\mathcal{I}$ *of T such that* $C^{\mathcal{I}} \neq 0 \rightarrow \mathcal{I}$ *is a model of C*

- **Concept equivalence**
  C equivalent to E wrt T iff $C^{\mathcal{I}} = E^{\mathcal{I}}$ *for every* $\mathcal{I}$ *of T* $\rightarrow$
  $C \equiv {}_{\mathcal{T}}E$ *or* $T \models C \equiv E$

- **Concept disjointness**
  C disjoint with E iff $C^{\mathcal{I}} \cap E^{\mathcal{I}} = \varnothing$

To add more expressiveness while retaining the computational power, various extensions of this $\mathcal{ALC}$ basic language were proposed in the literature by associating features such as:

- $\mathcal{S}$ transitive role (the roles can be transitive)
- $\mathcal{I}$ inverse role (e.g. "isPartOf"-"hasPart")
- $\mathcal{F}$ functional role (each individual has at most one relation with to an individual from the relation range)
- $\mathcal{H}$ role hierarchy (roles are in a subsumption hierarchy e.g. $hasSmallSize \sqsubseteq hasSize$)
- $\mathcal{O}$ nominals (name individuals appear in concepts e.g. $SlideID \equiv \{B10\}$)
- $\mathcal{N}$ number/cardinality constraints (the cardinality of a relation is restricted to "no less than" (e.g. ≥ 10 *hasNucleus* for expressing Tubule is formed by ten or more nucleus) and "no more than" (e.g. ≤ 2 *hasString.Nuclei* for expressing Tubule has no more than 2 strings of cell nuclei)
- $\mathcal{E}$ enumeration sets (concepts are enumerated in a finite set e.g. $SlideID \equiv \{B10, C10, B11\}$); they can be used in conjunction with nominals
- $\mathcal{D}$ data types (various data types such as integer, float, string)

Some examples: $\mathcal{SHOIN}(\mathcal{D})$ (the $\mathcal{ALC}$ extended with role transitivity, role hierarchy, nominals, role inverse, number restriction and data types), $\mathcal{SHIF}(\mathcal{D})$ (the $\mathcal{ALC}$ extended with role transitivity, role hierarchy, inverse roles, functional roles and data types), $\mathcal{ALCOIN}$ ($\mathcal{ALC}$ extended with nominals, inverse roles and cardinality constraints), etc.

From a temporal standpoint, another extension of DL reasoning was discussed in the literature to challenge the time dependency constraints in the reasoning [Artale and Franconi, 2001], [Artale et al., 2008], [Maris, 2008]. Similarly, probabilistic

extensions of DL tackled the probability issue [Lukasiewicz, 2007]. However, we do not go further into details, as these directions are not of current interest for us. An excellent overview and the most comprehensive presentation of DL theoretical foundation along with implementation in applications are given in [Baader et al., 2007].

2.2.2. Ontologies and Ontology Web Language

There are various definitions of an ontology. Merriam Webster's dictionary defines ontology as:

Definition 2.3. *Ontology is a branch of philosophy which studies what exists in all areas of reality*.

From another perspective which we also adopt, ontology is defined as following:

Definition 2.4. Ontology is an explicit specification of a conceptualization [Gruber, 1993].

According to [Chandrasekaran et al., 1999] theories in Artificial Intelligence fall into two broad categories: *mechanism* theories and *content* theories. They define ontologies as content theories about the types of objects, properties of objects and relations between objects that are possible in a specified domain of knowledge".
Ontologies consist of components called concepts, attributes, relations and instances:

- *concepts* or *classes* correspond to objects to be organized (e.g. projects, people, products, etc.);
- *attributes (slots*) are the traits of the objects (e.g. size, shape, etc.); and
- *relations* connect two objects or an object and a property to each other (e.g. « Patient » can be linked through the property « hasDisease » to a « Disease »);
- *instances* are the actual data in a given information system (e.g. "Slide NB50752007").

The set of classes represent the most important component of the ontology. Classes in ontologies describe sets of similar individuals in a certain domain. These classes can be divided into subclasses. Attributes give the description for the properties of the classes. The combination of classes and the instances (of its classes) creates the knowledge base.

Ontological analysis clarifies the structure of knowledge [Chandrasekaran et al., 1999]. Given a domain, its ontology forms the heart of any system of knowledge representation for that domain. Without ontologies, or the conceptualizations that underline knowledge, there cannot be a vocabulary for representing knowledge. Thus, the first step in devising an effective knowledge-representation system and vocabulary is to perform an effective ontological analysis of the field or domain. Weak analyses lead to incoherent knowledge bases.
As ontologies help to increase the efficiency and consistency of describing resources, they allow for more sophisticated functionalities in knowledge management and information retrieval for application development.

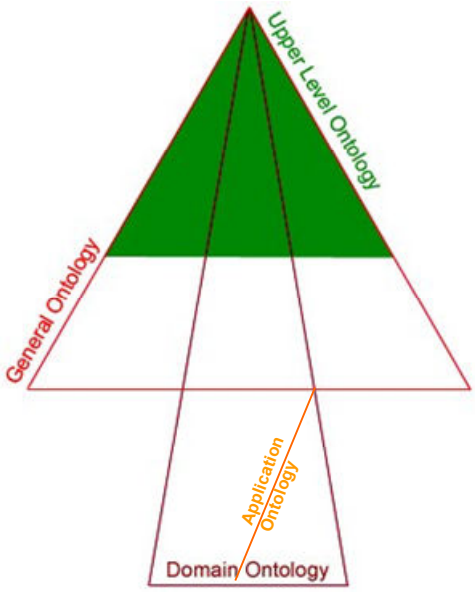


Figure 2.6. Types of ontologies

Ontologies may be categorized according to the domain they represent or the level of detail they provide [Bodenreider and Burgun, 2005], (Figure 2.6).

General ontologies represent knowledge at an intermediate level of detail independently of a specific task, whereas the ***upper level ontologies*** reflect theories of time and space, for example, and provide notions to which all concepts in existing ontologies are necessarily related.
Domain ontologies or reference ontologies represent knowledge about a particular part of the world, such as medicine, and should reflect the underlying reality through a theory of the domain represented. Within domain ontologies, the ontologies designed for specific tasks are called ***application ontologies***. Conversely, *reference ontologies* are developed independently of any particular purpose and serve as modules sharable across domains.

Table 2.8 gives a comparative analysis on reference ontologies and application ontologies based on [Menzel, 2007]. The characteristics they have shape our interest in the approach we propose in the following chapter.

REFERENCE ONTOLOGIES (RO)	APPLICATION ONTOLOGIES (AO)
Theoretical Focus: Representation	**Theoretical Focus: Reasoning**
• no computational efficiency concerns • takes advantage of full first-order logic language: ▪ arbitrary *n*-place predicates ▪ full classical negation ▪ unbounded, arbitrarily nested quantifiers	• computational applications • expressed in a computational language rooted in full first-order logic • reasoning classes and properties: ▪ unary and binary predicates ▪ conjunction & disjunction, but not negation ▪ limited quantification
Philosophical inclination: realism	**Philosophical inclination: pragmatism**
• *metaphysical realism* theWorld exists objectively in itself, independent of any mind • *epistemological realism* theWorld is knowable by us	• *metaphysical presumption* the falsity, irrelevance of metaphysical realism • *epistemological presumption* the underlying reality is probably unknowable anyway
Methodological emphasis on Truth	**Methodological emphasis on Fidelity**
• central *function* of an ontology : to represent theWorld accurately and comprehensively • *quality* of an ontology implies a function of its accuracy and comprehensiveness	• a *faithful expression* of the concepts of relevant domain experts or sources. • the quality of the ontology is determined entirely by the extent of its fidelity

Table 2.8. Reference versus application ontologies

Also stated in [Bodenreider and Burgun, 2005] is that core categories should be sharable across ontologies. Lower levels of upper level ontologies as well as general categories should be compatible with the equivalent semantic areas in the corresponding domain ontologies. For example, *Disease* in a general ontology should be compatible with that concept in a biomedical ontology.
In addition, generic theories and meta-level categories should be shared by every type, in every kind of ontology. For example, a representation of anatomy should re-use a generic theory of spatial objects. In turn, as anatomy is central to biomedicine and essentially stable, an ontology of anatomy can serve as a reference for ontologies relying on a representation of the human body, e.g., for an ontology of Diseases.
Inspired by the decidability power of DL, the World Wide Web Consortium (W3C) proposed the creation of a web ontology standard language capable of both expressivity and computability [Horrocks et al., 2003]. Since its introduction as OWL 1.0 [Horrocks et al., 2003] and [Horrocks et al., 2007], practical experiences have been encouraging. However, as a first attempt it revealed the need for improvements in some areas and hence OWL 2.0 was recently launched [Grau et al., 2008]. Regardless of the version, OWL1[1] standard comes with three increasingly expressive sublanguages based on the family of $\mathcal{SH}$ and designed for different types of usage:

- ***OWL Lite*** *(corresponds to* $\mathcal{SHIF}$ **(**$\mathcal{D}$**)**) supports cardinality constraints, yet it only permits cardinality values of 0 or 1. It is efficiently used for classification hierarchy and simple constraint features [Artale et al., 2007].

[1] "The natural acronym for *Web Ontology Language* is *WOL* instead of *OWL*. However, OWL was proposed as an easily pronounced acronym that would yield good logos, suggest wisdom, and honor William A. Martin's *One World Language* [*Knowledge Representation]* project from the 1970s. And, to quote Guus Schreiber, <<Why not be inconsistent in at least one aspect of a language which is all about consistency?>>" [Obitko, 2007]

- ***OWL DL*** *(corresponds to* $\mathcal{SHOIN}$ **($\mathcal{D}$))** offers full expressivity combined with computational completeness and decidability power. One constraint imposed by this language is the type distinction; a class can not be an individual or a property, similarly a property can not be a class or individual. It is serialized using RDF/XML syntax, where the syntax is viewed as RDF graph, composed of RDF triples sets (subject- predicate - object) and the semantics is an extension of RDF semantics.
- ***OWL Full*** offers maximum expressiveness but with no computational warrantees. For example, the data type property can be considered as inverse functional property, and a class can be treated in the same time as a set of individuals and as an individual by itself. From the perspective of serialization, OWL Full has the syntactic freedom of RDF, hence the reasoning does not offer a full support for every characteristic of this language.

Based on this classification, we adopt OWL DL as our formalism language. Thus we will only discuss this language further on.
Table 2.9 and Table 2.10 show the OWL-DL syntax and semantics for creating the vocabulary elements of ontology, as presented in [Obitko, 2007].

Abstract Syntax	DL Syntax	Semantics
Descriptions (C)		
A (URI Reference)	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
owl:Thing	$\top$	$\texttt{owl:Thing}^{\mathcal{I}} = \Delta^{\mathcal{I}}$
owl:Nothing	$\bot$	$\texttt{owl:Nothing}^{\mathcal{I}} = \emptyset$
intersectionOf($C_1\ C_2 \ldots$)	$C_1 \sqcap C_2$	$C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$
unionOf($C_1\ C_2\ \ldots$)	$C_1 \sqcup C_2$	$C_1^{\mathcal{I}} \cup C_2^{\mathcal{I}}$
complementOf(C)	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
oneOf($o_1 \ldots$)	$\{o_1, \ldots\}$	$\{o_1^{\mathcal{I}}, \ldots\}$
restriction(R someValuesFrom(C))	$\exists R.C$	$\{x \mid \exists y\ \ (x,y) \in R^{\mathcal{I}} \cup y \in C^{\mathcal{I}}\}$
restriction(R allValuesFrom(C))	$\forall R.C$	$\{x \mid \forall y\ \ (x,y) \in R^{\mathcal{I}} \rightarrow y \in C^{\mathcal{I}}\}$
restriction(R hasValue(o))	$R : o$	$\{x \mid (x, o^{\mathcal{I}}) \in R^{\mathcal{I}}\}$
restriction(R minCardinality(n))	$\geqslant nR$	$\{a \in \Delta^{\mathcal{I}} \mid\ \mid\{b \mid (a,b) \in R^{\mathcal{I}}\}\mid \geq n\}$
restriction(R maxCardinality(n))	$\leqslant nR$	$\{a \in \Delta^{\mathcal{I}} \mid\ \mid\{b \mid (a,b) \in R^{\mathcal{I}}\}\mid \leq n\}$
restriction(U someValuesFrom(D))	$\exists U.D$	$\{x \mid \exists y\ \ (x,y) \in U^{\mathcal{I}} \cup y \in D^{\mathcal{D}}\}$
restriction(U allValuesFrom(D))	$\forall U.D$	$\{x \mid \forall y\ \ (x,y) \in U^{\mathcal{I}} \rightarrow y \in D^{\mathcal{D}}\}$
restriction(U hasValue(v))	$U : v$	$\{x \mid (x, v^{\mathcal{I}}) \in U^{\mathcal{I}}\}$
restriction(U minCardinality(n))	$\geqslant nU$	$\{a \in \Delta^{\mathcal{I}} \mid\ \mid\{b \mid (a,b) \in U^{\mathcal{I}}\}\mid \geq n\}$
restriction(U maxCardinality(n))	$\leqslant nU$	$\{a \in \Delta^{\mathcal{I}} \mid\ \mid\{b \mid (a,b) \in U^{\mathcal{I}}\}\mid \leq n\}$
Data Ranges (D)		
D (URI reference)	D	$D^{\mathcal{D}} \subseteq \Delta_{\mathcal{D}}^{\mathcal{I}}$
oneOf($v_1 \ldots,$)	$\{v_1 \ldots,\}$	$\{v_1^{\mathcal{I}} \ldots,\}$
Object Properties (R)		
R (URI reference)	R	$\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
	R^-	$(R^{\mathcal{I}})^-$
Datatype Properties (U)		
U (URI reference)	U	$U^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta_{\mathcal{D}}^{\mathcal{I}}$
Individuals (o)		
o (URI reference)	o	$o^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
Data Values (v)		
v (RDF literal)	v	$v^{\mathcal{D}}$

Table 2.9. OWL- DL syntax and semantics [Obitko, 2007]

The description of the language constructs is given in detail in [McGuiness and Harmelen, 2009]. At this point, there are some remarks to be considered. The first one is related to the correspondence of DL terminology to OWL terminology.

DL	*OWL*
concept	→ *class*
role	→ *property*
individual	→ *individual / object*

Abstract Syntax	DL Syntax	Semantics
Classes		
Class(A partial $C_1 \dots C_n$)	$A \sqsubseteq C_1 \sqcap \dots \sqcap C_n$	$A^{\mathcal{I}} \subseteq C_1^{\mathcal{I}} \cap \dots \cap C_n^{\mathcal{I}}$
Class(A complete $C_1 \dots C_n$)	$A \equiv C_1 \sqcap \dots \sqcap C_n$	$A^{\mathcal{I}} = C_1^{\mathcal{I}} \cap \dots \cap C_n^{\mathcal{I}}$
EnuneratedClass(A $o_1 \dots o_n$)	$A \equiv \{o_1, \dots, o_n\}$	$A^{\mathcal{I}} = \{o_1^{\mathcal{I}}, \dots, o_n^{\mathcal{I}}\}$
SubClassOf(C_1 C_2)	$C_1 \sqsubseteq C_2$	$C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$
EquivalentClasses($C_1 \dots C_n$)	$C_1 \equiv \dots \equiv C_n$	$C_1^{\mathcal{I}} = \dots = C_n^{\mathcal{I}}$
DisjointClasses($C_1 \dots C_n$)	$C_i \sqcap C_j = \bot, i \neq j$	$C_i^{\mathcal{I}} \cap C_j^{\mathcal{I}} = \emptyset, i \neq j$
Datatype(D)		$D^{\mathcal{C}} \Delta_D^{\mathcal{I}}$
Datatype Properties		
DatatypeProperty(		
U super(U_1) ... super(U_n)	$U \sqsubseteq U_i$	$U^{\mathcal{I}} \subseteq U_i^{\mathcal{I}}$
domain(C_1) ... domain(C_m)	$\geqslant 1\, U \sqsubseteq C_i$	$U^{\mathcal{I}} \subseteq C_i^{\mathcal{I}} \times \Delta_D^{\mathcal{I}}$
range(D_1) ... range(D_l)	$\top \sqsubseteq \forall U.D_i$	$U^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times D_i^{\mathcal{I}}$
[Functional])	$\top \sqsubseteq \leqslant 1 U$	U_i is functional
SubPropertyOf(U_1 U_2)	$U_1 \sqsubseteq U_2$	$U_1^{\mathcal{I}} \subseteq U_2^{\mathcal{I}}$
EquivalentProperties($U_1 \dots U_n$)	$U_1 \equiv \dots \equiv U_n$	$U_1^{\mathcal{I}} = \dots = U_n^{\mathcal{I}}$
Object Properties		
ObjectProperty(		
R super(R_1) ... super(R_n)	$R \sqsubseteq R_i$	$R^{\mathcal{I}} \subseteq R_i^{\mathcal{I}}$
domain(C_1) ... domain(C_m)	$\geqslant 1\, R \sqsubseteq C_i$	$R^{\mathcal{I}} \subseteq C_i^{\mathcal{I}} \times \Delta_D^{\mathcal{I}}$
range(C_1) ... range(C_l)	$\top \sqsubseteq \forall R.C_i$	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times C_i^{\mathcal{I}}$
[inverseOf (R_0)]	$R \equiv (R_0^-)$	$R^{\mathcal{I}} = (R_0^{\mathcal{I}})^-$
[Symmetric]	$R \equiv (R^-)$	$R^{\mathcal{I}} = (R^{\mathcal{I}})^-$
[Functional]	$\top \sqsubseteq \leqslant 1 R$	$R^{\mathcal{I}}$ is functional
[InverseFunctional]	$\top \sqsubseteq \leqslant 1 R^-$	$(R^{\mathcal{I}})^-$ is functional
[Transitive])	$Tr(R)$	$R^{\mathcal{I}} = (R^{\mathcal{I}})^+$
SubPropertyOf(R_1 R_2)	$R_1 \sqsubseteq R_2$	$R_1^{\mathcal{I}} \subseteq R_2^{\mathcal{I}}$
EquivalentProperties($R_1 \dots R_n$)	$R_1 \equiv \dots \equiv R_n$	$R_1^{\mathcal{I}} = \dots = R_n^{\mathcal{I}}$
Annotation		
AnnotationProperty(S)		
Individuals		
Individual(		
o type(C_1) ... type(C_n)	$o \in C_i$	$o^{\mathcal{I}} \in C_i^{\mathcal{I}}$
value(R_1 o_1) ... value(R_n o_n)	$\{o, o_i\} \in R_i$	$\{o^{\mathcal{I}}, o_i^{\mathcal{I}}\} \in R_i^{\mathcal{I}}$
value(U_1 v_1) ... value(U_n v_n))	$\{o, v_i\} \in U_i$	$\{o^{\mathcal{I}}, v_i^{\mathcal{I}}\} \in U_i^{\mathcal{I}}$
SameIndividual($o_1 \dots o_n$)	$o_1 = \dots = o_n$	$o_1^{\mathcal{I}} = \dots = o_n^{\mathcal{I}}$
DifferentIndividual($o_1 \dots o_n$)	$o_i \neq o_j, i \neq j$	$o_i^{\mathcal{I}} \neq o_j^{\mathcal{I}}, i \neq j$

Table 2.10.OWL- DL axioms and facts [Obitko, 2007]

The second remark is with respect to ontology versus knowledge-base. The question that arises is: what is the difference between ontology and knowledge-base?

Ontology defines the structure of stored data such as what types of entities are recorded, what their relationships are, what are the logical operators used to build up the description of entities and properties. Mostly important, ontologies s*upport reasoning about knowledge stored in the knowledge-base.* The latter stores knowledge in a computer readable format, thus it has automated reasoning. The set of data is contained in the knowledge-base in the form of rules that describe the knowledge in a logically consistent manner.

2.2.3. Rules. Semantic Web Rule Language

Rules are needed for several reasons [Artale et al., 2007]: the existing rule set can be reused, higher expressivity can be added to OWL and sometimes it is easier to read and write rules with a rule language. However there are also some limitations on the decidability power when there is high expressivity. We will discuss this aspect later.
SWRL is a combination of OWL with RuleML (a rule language) proposed to benefit of the advantages of ontologies and rules in the same framework [Boley et al., 2005]. RuleML represents the Datalog sublanguage of Horn clause [Horrocks et al., 2004], whilst SWRL is a restricted version of RuleML. One constraint is that it allows only unary/binary relations, which implies that n-ary relations are transformed to binary relations.
A comprehensive presentation of SWRL in given in [Karimi et al., 2008], therefore we present the syntax, semantics and the definitions for safety rules accordingly.

SWRL syntax

SWRL atoms are defined as follows:

$$Atom \leftarrow C(i) \mid D(v) \mid R(i,j)U(i,v) \mid builtIn(p,v_1,...,v_n) \mid i = j \mid i \neq j \quad (2.15)$$

where C is a class, D is data type, R is object property, U is data type property I, j are object variable names or object individual names, $v_1, \ldots v_n$ are data type variable names or data type value names, p is built-in name.
A SWRL rule syntax is of the form:

$$a \leftarrow b_1,...b_n \quad (2.16)$$

Where a is the head, or the consequent (an atom), and b stands for the body, the precedent (all atoms).

SWRL semantics

Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \Delta^{D}, .^{\mathcal{I}}, .^{D})$ (2.17)

where $\mathcal{I}$ is the interpretation, $\Delta^{\mathcal{I}}$ is the object interpretation domain, Δ^{D} - datatype interpretation domain, $.^{\mathcal{I}}$ object interpretation function, $.^{D}$ -datatype interpretation function, and $\Delta^{\mathcal{I}} \cap \Delta^{D} = \varnothing$, such that :

$$V_{\mathcal{I}X} \rightarrow P(\Delta^{\mathcal{I}}) \qquad V_{DX} \rightarrow P(\Delta^{D}) \tag{2.18}$$

where $V_{\mathcal{I}X}$ denotes the object variables, V_{DX} the datatype variables, and P the power-set operator.

SWRL atoms	**Semantics**
$C(a)$	$i^{\mathcal{I}} \in C^{\mathcal{I}}$
$R(i,j)$	$(i^{\mathcal{I}}, j^{\mathcal{I}}) \in R^{\mathcal{I}}$
$U(a,v)$	$(a^{\mathcal{I}}, v^{D}) \in U^{\mathcal{I}}$
$D(v)$	$v^{D} \in \Delta^{D}$
$builtIn(p, v_1 \ldots v_n)$	$(v_1{}^{D}, \ldots v_n{}^{D} \in p^{D})$
$a = b \,/\, a \neq b$	$a^{\mathcal{I}} = b^{\mathcal{I}} \,/\, a^{\mathcal{I}} \neq b^{\mathcal{I}}$

Table 2.11. SWRL semantics with bindings B($\mathcal{I}$) [Karimi, 2008]

SWRL atoms in the antecedent are satisfied if either the antecedent is empty or if every atom is satisfied. SWRL atom from the consequent is satisfied if the consequent is non-empty and it is satisfied.
A rule is satisfied by an interpretation $\mathcal{I}$ iff every binding B($\mathcal{I}$) satisfies the antecedent and B($\mathcal{I}$) satisfies the consequent.

SWRL Knowledge-base

A SWRL knowledge-base $\mathcal{K}$ is defined as follows:

$$K = (\sum, P) \tag{2.19}$$

where $\sum$ is the $\mathcal{SHOIN}(\mathcal{D})$ knowledge-base and P a finite set of rules.

When some rules from SWRL can be easily expressed into DL without the help of SWRL constructs, these rules are called syntactic sugar. However, it is not always easy to translate a SWRL rule to DL, for it depends on the number of shared variables between the antecedent and the consequent. The translation is possible in the following conditions:

1. if zero variables are shared and/or one individual is shared
2. if one variable is shared

The consequent and antecedent become two conjunctive queries. The conjunctive terms form a directed graph, where each node is a variable or a named individual and each edge is a relation.
The resulting queries graphs are translated into class expressions, using rolling-up technique. Each outgoing edge is represented as existential quantifier and edges are

presented as restrictions. Each outgoing edge of (?x, ?y) :R transforms as $\exists R.Y$, where Y is the named class restriction on variable ?y.
After that, the antecedent becomes the subclass of the consequent.

Although there are rules that can be translated from SWRL to DL, SWRL can express rules DL cannot. Yet there is a tradeoff in terms of decidability.
To overcome this issue, SWRL DL safe rules were proposed. The safety conditions to SWRL imply adding additional expressive power and maintaining computational power. As both OWL and Datalog are decidable, the desideratum is to make their combination decidable too.
A datalog rule is safe if every variable in the consequent appears in the antecedent. The DL safety rules are created with more restrictions. The definitions for strong and weak DL safety are given in [Parsia et al., 2005], [Karimi, 2008].

Definition 2.5. **Strong DL safety**
Let be $\sum$ an OWL-DL ontology and P a Datalog program. A rule *r* in *P* is **strongly DL-safe** if each variable in *r* occurs in a non-DL atom *in the rule body*. The program *P* is strongly DL-safe if all its rules are strongly DL-safe.

Definition 2.6. **Weak DL safety**
Let be $\sum$ an OWL-DL ontology and P a Datalog program. A rule *r* in *P* is **weakly DL-safe** if each variable in *r* occurs in a non-DL atom *in the rule*. The program *P* is weakly DL-safe if all its rules are weakly DL-safe.

Definition 2.7. **Role safety**
Let be $\sum$ an OWL-DL ontology and *P* a Datalog program. A rule *r* in *P* is **role safe** if for each DL atom p with arity 2 in the antecedent of r at least one variable in p appears in a non-DL atom q in the antecedent of r and q never appears in the consequent of any rule in the program *P*.

Various approaches of combining OWL and SWRL have been proposed. In [Samuel et al., 2008], two situations are investigated in a system called SWORIER developed in Prolog: using OWL with SWRL to develop an integrative ontology/rule language and layering rules on top of ontology with RuleML and OWL. Also for Prolog, [Herchenröder, 2006] proposes an extension of tableau-based algorithm of DL for a lightweight semantic web.

2.3. Spatial Representation

If the main goal of the previous two sections was to give a theoretical foundation of knowledge representation with respect to image and concepts, the purpose of this section is to present one particular kind of knowledge representation that is relevant from both perspectives: the spatial representation.
Spatial structures and relations are essential to the perception of the world (such as surrounded by, shape, relative position, location, etc) and the ways we derive consequences from them. While the moving from space to theories of space cannot be complete, there is a plethora of fields of research in spatial representation and reasoning, ranging from philosophical to AI vision and robotics communities,

geographic information science, semantics representation in the study of natural language, modal logics, to spatial representation in medical science. Consequently, the variety of theories speaks for itself. An extensive state-of-the-art literature in spatial representation theory and practice is given in [Aiello, 2002]. Another overview is presented in the work of [Cohn and Renz, 2008]. Both cover a vast range of approaches in theories and practical implementations.
One of the reasons for such wideness is the complexity of the space itself and the fact that a representation must reflect the spatial structures and relations from what it is relevant in its view.
To this end, we limit the direction of research in this thesis to one particular kind of spatial representation and even for this, we refine the discussion to only the approaches relevant to the context of our work so to derive the fundamental mindset for our approach.
Having said that, and in light of the convergence between image and semantics, the following discussion evolves around the *quantitative representation versus qualitative representation of space*.
Quantitative representation is viewed mostly as the traditional way of representing the information from image.
The quantitative representation is defined as a representation based on metric measurements, coordinates and numerical values. This kind of approach is generally applied in robotics, computational geometry, image processing algorithms. To take a concrete medical example: the description of the size of a tumor in a histopathological image.
However, this approach has several drawbacks. [Brageul and Guesgen, 2007] gives a synthesis of the most important limitations in dealing with quantitative representations:

- if no exact or precise data is available, the quantitative representation is not reliable.
- *the positions of all the objects have to be known, regardless of whether we need them or not.*
- *when a value is not known exactly, it has to be either ignored or assigned*. Hence, when dealing with partial and uncertain information, a quantitative representation may not produce reliable results in the process of inference.
- missing adequacy which means to *use quantities regardless of the nature of facts, whether qualitative or quantitative.*
- we resemble these limitations to what is called the "aquarium metaphor" explained in [Freksa, 1991].

To grapple with these specific problems the paradigm of qualitative representation gain a lot of interest based on the intrinsic properties it has.

A qualitative representation is defined as a representation that uses symbolic knowledge. In a qualitative representation, there is no need to know the positions of all the objects. It is sufficient to represent as much spatial information as needed.
For instance, it is sufficient to know that a *large size* of a nucleus represents a high score for the assessment of the tumor and encode it as *large size*. Other aspects depend on the depth, the granularity of the representation. If an object is close to the periphery of the tumor in the tissue, it is not necessary to describe in detail what periphery means. It suffices to say that the periphery is known as opposed to the central zone for this particular kind of information. This goes hand in hand with

the situation when no precise quantitative data is available. In such context, a qualitative representation is highly desirable.

There is also a discussion around the relation between quantitative and qualitative representation and reasoning, as to find whether they depend one on the other or not. [Kadijevic, 2002] investigates this problem by testing both paradigms on mathematical abilities to solve proportional problems. The assumption that qualitative representation should precede the quantitative representation for it would provide better quantitative thinking did not hold. However, the study evidences that qualitative reasoning is more efficient than traditional quantitative reasoning.

One of the fundamental assumptions of qualitative spatial representation is that spatial situations from the real world are represented by defining spatial relationships between the given entities.
Formally, a *relation* R is a set of tuples $(d_1,\ldots,d_k)$ of the same arity k, where d_i is a member of a corresponding *domain* D_i. In other words, a relation R of arity k is a subset of the cross-product of k domains [Cohn and Renz, 2008]:

$$R \subseteq D_1 \text{ x} \ldots \text{x} D_k \tag{2.20}$$

Spatial relations are generally considered *binary relations* and consequently the considered domains are viewed as identical, namely, the set of all spatial entities of a particular space. Hence, according to [Cohn and Renz, 2008], in these cases spatial relations are of the form:

$$R = \{(a,b) \mid a,b \in D\}. \tag{2.21}$$

Given a set of relations $R = \{R_1, \ldots Rn\}$ algebraic operators such as union, intersection, complement, converse, or composition of relations can be used in order to obtain an *algebra of relations*. The basic algebra relations commonly found in qualitative reasoning are the so called jointly exhaustive and pair-wise disjoint (JEPD), which essentially means that the relationship between two spatial entities must be exactly one of the JEPD relation, each tuple is a member of exactly one relation.

Various set of spatial JEPD relations, combination of them have been discussed in the specialized literature and due to the complexity of space, variety of theories of space aroused. There are mereological theories of parts and wholes [Donnelli et al., 2005], topological theories with objects connections and limit points (disjointness, inclusion), theories for orientation and distances. We provide the definition of mereology as we will discuss about it, individually or in connection with others, in the next chapters.

Definition 2.8. Mereology is the theory of parthood relations: of the relations of part to whole and of the relations of part to part within a whole [Varzi, 2009].
In the same reference there is a discussion on the similarities and differences between mereology and set theory from which mereology evolved (for instance, mereology is different to set theory in that is committed to the existence of neither the *abstracta- the whole can be as concrete as the parts, nor concreta - the parts*

can be as abstract as the whole, but mereology and set theory are similar in that both are attempts to lay down the general principles underlying the relationships between an entity and its constituent parts/ a set and its members.

None of the above approaches (mereology, topology, etc) can provide a calculus to represent and reasoning with all aspects of the space. A combination of them would be more appropriate if the domain that we want to represent deals with a combination of spatial relations.

Consequently, mereo-topological approach handles parthood and external connections of objects [Cohn and Renz, 2008], topology, orientation and distance are combined in qualitative framework in application of geometric informational space and robot navigation [Brageul and Guesgen, 2007], closeness and distance relations via a method oriented on propositional dynamic logic [Burrieza et al., 2009]. Nevertheless, there are variants of representing spatial knowledge through a multitude of methods. For instance, methods as 4-intersection model, RCC-8 (region conection calculus) was proposed for topologic representation, double-cross calculus or dipole-relation algebra for orientation representation, or composition tables for topology, orientation and distances [Brageul and Guesgen, 2007], [Cohn and Renz, 2008].

We do not go further into detail; for reference, a comprehensive overview of them is given in [Cohn and Renz, 2008]. We will come to some relations and their characteristics in chapter 5 where we discuss the spatial representation we advance in our approach.

2.4. Conclusions

This chapter was devoted to laying the foundation of knowledge representation and reasoning from different perspectives. We introduced the concept of knowledge representation first and noted that representation and reasoning are very strongly entangled.
We then discussed two paradigms for image representation, the CBIR and CBR namely, moving along to semantic representation approaches. At this point we introduced a logic formalism, the DL, which proved to be very useful in creating the language OWL for the semantic web, so necessarily to make the representation understandable by both humans and machines. As we deal with structures and rules on daily basis representation, a formal language for supporting rules combined with ontologies was presented.
To consider both CBIR and CBR as methodologies (meaning generic) represents an asset of our thesis. It implies that they can be further mapped with any kind of application desired and it also allows us to provide a pertinent comparative analysis on each process they involve.
Regarding semantic representation, one could say that a framework in which DL and OWL work together is an efficient solution for achieving high expressivity and computational power in the same time. When rules are needed, the tradeoff between high expressivity and decidability is present but there are answers to keep both as we discuss in chapter 6.

Our reasoning for providing such an approach for knowledge representation is driven by the main goal of this thesis, which is to propose a framework to connect representation of concepts with representation of image features. In this light, the natural demarche was to study the spatial representation and reasoning theory and practice. The study was based upon the grounds of a qualitative spatial representation, after discussing the advantages it has against the quantitative paradigm.

We mention once again that we did not go into the very details of each aspect presented. For instance, the characteristics and problems of OWL and SWRL will be approached in a more granular manner in chapter 6 which regards the implementation of the model we propose in chapter 4. Additionally we did not give examples of OWL or SWRL applications, for the range is too wide and not relevant to our direction of research. We discuss their applications with respect to medical field. Even for CBIR and CBR analysis from indexing, retrieval and refining perspectives, we limited the study to emphasize the characteristics of each axis so to set up the basis of our approach. Also, the spatial details are discussed along the spatial relations we refer to in our proposal.

To sum up, the assumptions that we contribute with, towards our approach are:

- image representation as knowledge representation
- CBIR and CBR as methodologies, not technologies
- comparative analysis of CBIR and CBR from indexing, retrieval and refining perspectives

3. Knowledge Representation and Reasoning in Medical Applications

Based on the study of various ways of doing knowledge representation and reasoning conducted in chapter 2, the purpose of this whole section is to analyze the approaches proposed in medical applications. In the first part we describe the implementation of CBIR and CBR and their specific issues with respect to medical field. We shift to biomedical ontologies and the role the spatial representation plays in such a formal representation. Final remarks are given in the last section of the chapter.

3.1. Case- Based Reasoning versus Content- Based Image Retrieval

As a general overview, both CBIR and CBR are positioned at the confluence of some related areas, revealing their cross-discipline orientation. These crossroads also show the domains from where they emerged and extended afterwards.

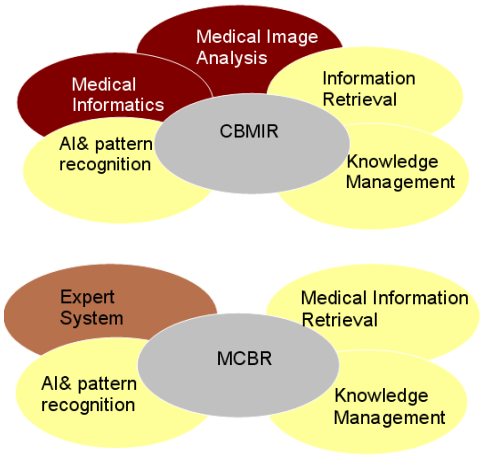


Figure 3.1. Content-Based Medical Image Retrieval (CBMIR) and Medical Case-Based Reasoning (MCBR) related fields

Figure 3.1 shows in contrast the interconnected domains for Content-Based Medical Image Retrieval (CBMIR) and Medical CBR (MCBR), respectively. The great success of CBIR and CBR witnessed in the scientific literature emphasized a potentially significant impact for diagnosis and prognosis assistance in medical communities.

At its origin, CBIR was proposed to meet the needs of the semantic web, as a more general approach in terms of information retrieval and knowledge management. Since its arrival, interactions within fields such as computer vision, machine learning, and data mining contributed to CBIR community expansion and development [Wang and Fartha, 2005].
Tracing back, the CBR roots are found in the Artificial Intelligence, as derived from Knowledge-Based Systems. CBR is positioned at the confluence of database systems and expert systems [Richter, 2003].
Within the last years, the number of digital images produced in the medical field is continuing to increase in large amounts. Hence a crucial need to design CBIR system to assist in the diagnosis, prognosis has been highlighted. A classification of CBIR medical applications, describing the most representative ones: PACS (Picture Achieving and Communication System) with the extended version cbPACS [Traina et al., 2005], IRMA [Lehman et al., 2006] and MedGift [Hidki et al., 2007], in an a posteriori approach is done in [Deserno et al., 2007]. However, the promising development on CBIR in the scientific community did not accrue in the same manner in the health science domain.
The reason for such a lack is attributed partially to various gaps of CBIR comprehensively described by [Smeulders et al., 2000], [Müller, 2004], [Deserno et al., 2007], [Datta et al., 2008].
Despite of some minor disagreements of the authors, a compilation of all gaps is given by Table 3.1, with our own emphasis on *perception gap* instead of *aesthetic gap* proposed by [Datta et al., 2008] We consider that the perception terminology for this gap issue is more appropriate to be taken in the medical field, rather than an aesthetic terminology.
Another important aspect concerns the semantic gap which is defined as:

Definition 3.1. Semantic gap **is** ***the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation*** [Smeulders et al., 2000].

Many papers considered the semantic gap the single issue in applying CBIR in medical domain (CBMIR). As stated above, there are, in fact, multiple gaps that hinder this usage and moreover, the semantic gap is somewhat dependent on the context of the application. Hence, a generic content gap which includes both semantic and context gaps is proposed in [Müller, 2004]. We will further focus on the content gap only.

Another shortcoming related to CBMIR is the relevance feedback lack. Since the discussion and proposal of relevance feedback integration in the scientific applications- where it also missed or it was weak until recently – it came naturally that this is the need of the hour for CBMIR too. Same implications of including relevance feedback in the CBIR process apply to CBMIR, especially the one related to the semantics, which is highly important in the medical knowledge-based domain. From the system's performance perspective, a relevance feedback could improve both precision and recall. In the approach taken by [Lamard et al., 2007], using signatures derived from the coefficients of the wavelet transform, the precision is showed to reach 79.5% for retinal databases, while the retrieval efficiency for face database and mammography data are presented to be very good.

CBMIR gaps	**Characteristics**
Content	modeling & understanding image/information-real image/information
Features	computational numerical features- real image/information
Performance	application, integration, indexing, evaluation
Usability	query, feedback, refinement
Perception	visual information perception - real image/information perception
Sensory	Information description – real image/information

Table 3.1. CBMIR gaps

Also related to wavelet transform, [Quellec at al., 2010c] proposes a method to adapt a multidimensional wavelet filter bank, based on a non-separable lifting scheme framework, to any specific problem.
The method is applied to CBIR, evaluated on two medical image databases (Digital Database for Screening Mammography-DDSM, Diabetic Retinopathy Database-DRD) and two non-medical image databases (Face Database-FD and Vision Texture Database-VisTex). The performances of the adapted wavelet filter bank over the non-adapted wavelet filter bank are studied and there are shown to be higher for every database. When compared to a similar CBIR system based on an adaptive separable wavelet transform, the performances of the non-separable wavelet based system are notably higher on three out of the four databases (DDSM, DRD and VisTex), and comparable on the other one (FD).
When a query expansion is involved, it is shown that relevance feedback improves the recall metric with a strategy of balancing the positive and the negative feedback (usually adopted where there is a relevance feedback, even if weak) [Manning et al., 2008].
At this point, to better understand the idea, an elaboration is to be done with respect to positive and negative feedback. The key point is: more feedback, better results. To achieve this aim, positive and negative feedback are employed, with focus on the positive feedback that turn out to be more valuable than the negative feedback. But the problem comes when, for instance, images containing a small number of features are returned, and hence there is too much negative feedback. A solution of separately weighting the positive the negative results, or an automated feedback and furthermore a probabilistic relevance feedback -inspired from the Rocchio method applied in information retrieval - is proposed in [Deselaers and Müller, 2005].

User interfaces are also a drawback of CBIR when it comes to apply them in health science domain. How to manage with medical multimedia data is, nevertheless, challenging, due to the need of visual search the physicians have to operate with. The user interface with which the clinician communicates implies an adaptation to the technology and confidence in the same time, of the medical doctor to be able to efficiently use it. Issues related to the user interfaces are given in addition, by the lack of a missing combined framework to perform the system evaluation. An attempt of this kind is found in [Müller et al., 2003].

In the same time, CBR systems are highly promising to be used in clinical practice mainly due to their cognitive adequateness (similar reasoning with the physicians) characteristic and the explicit experience involved (the property of CBR to adjust itself to meet the need of a specific medical doctor or hospital). Their baseline principle is more closed to the medical reasoning; therefore CBR systems have been designed for diagnostics, tutoring and planning. Yet, issues such unreliability, adaptation or concentration of reference are still facts to face with in medical CBR.
CBR systems become more reliable with the percentage of domain knowledge covered in its development. Still, the reliability cannot be fully guaranteed [Bichindaritz, 2003]. We consider that a solution to this issue is intrinsic given by the duality of objective and subjective knowledge advantage of CBR, since the knowledge can come from different parts. Thus it is possible in the first hand to complete the domain knowledge, and secondly, to balance the system through the various information provided by different parties.

How to deal with the case adaptation when a multitude of features are involved represents another challenge. Due to this strong issue, some approaches do not even have the case adaptation phase, stopping at the retrieval phase [Perner, 2001]. But in our opinion, avoiding the adaptation problem makes CBR similar with CBIR in the sense that it will work more likely as a CBIR rather than a standard CBR.
Although a generalization (abstraction) of features or a more efficient identification of features could solve this problem to some degree, there are, nevertheless, trade-offs. To support this idea, constraints in terms of reducing the set of solutions by checking the contraindications or contradictions were also proposed as an interesting solution to this problem; but it can be applied only for very specific situation.
Another idea is to use adaptation rules, but although the technique is considered to be general, the rules have to be specifically oriented to the application. A presentation of all these solutions is met in [Schmidt et al., 2003].

Related to the case adaptation, concentration on reference refers to the fact that a CBR system is limited in its functionality if a suitable case does not exist in the case repository. In other words it could be prone to inexactness when the current solution is not identical with the previous one. Hence, the outcome of results (prognosis or diagnosis) could be influenced and may vary in time. To this end, the third Res -the revise step- is to be taken into account. This input is given by the case adaptation's suggested solution and evaluated in the case verification phase. However, little work has been done with respect to this step. We envision that probabilistic methods could be a modality to provide this step in order to achieve the final task of the CBR.

A detailed classification of various influential medical CBR systems of last years, from the purpose-oriented and construction-oriented perspectives is given in [Nillson and Sollenborn, 2004], Table 3.2.

Medical CBR paradigms [Nillson and Sollenborn, 2004],	
Purpose-oriented systems	➢ diagnosis ➢ classification ➢ tutoring ➢ planning
Construction-oriented systems	➢ hybrid ➢ adaptive ➢ autonomicity ➢ constraints

Table 3.2. Medical CBR paradigms

	Advantages	**Drawbacks**
Medical CBIR [Müller, 2004]	➢ increasing rate of image production ➢ applications in diagnosis, teaching & research	➢ relevance feedback ➢ user interfaces ➢ performance
Medical CBR [Nillson and Sollenborn, 2004] [Schmidt and Gierl, 2001], [Holt et al., 2006]	➢ cognitive adequateness ➢ explicit experience ➢ duality of objective & subjective knowledge ➢ system integration ➢ application in diagnosis, teaching & research	➢ adaptation ➢ unreliability ➢ concentration on reference

Table 3.3. CBIR & CBR in medical field

Table 3.3 summarizes main advantages and drawbacks of CBIR and CBR in medical field discussed above. A presentation on the current work in CBR and the future of CBR in medical applications is provided by [Bichindaritz and Marling, 2006].

3.2. From CBIR and CBR to Formal Representation and Reasoning. Biomedical Ontologies

Having studied the concepts of CBIR and CBR, their characteristics in terms of indexing, retrieval and refining followed by a synthesis on their application in the medical field, the question that arouse at this point of research was this: by what means to connect CBIR and CBR in order to achieve a refined representation and thus reliable diagnosis or prognosis assistance from the medical standpoint?

At this point we came to realize two major facts:

Reasoning is the hallmark of both CBIR and CBR. Furthermore, reasoning is the support for a reliable semantic annotation and retrieval.

These two conclusions are strongly related. Apart from which axis we analyze, that is indexing or retrieval or refining, the element that drives them is the reasoning. CBIR is an image-based reasoning while the CBR is a text or semantic-based reasoning. The ability to use the indexed knowledge in the process of retrieval and to refine the results to have a more accurate representation and thus reliable output, that is the characteristic of reasoning. In the medical field, the reasoning is even more visible since it resembles with the reasoning of physicians.
Hence, in order to obtain a refined representation we need reasoning. Following that, another question that arouse was: what kind of representation to vote for in order to allow semantic annotation and retrieval? Should we go for a combined CBIR – CBR for medical applications, since our desideratum is to make progress in medical field, if possible by keeping the advantages of both techniques? In another medical context, the protein crystallization more specifically, the authors proposed an integration of CBIR into a CBR framework [Jurisica et al., 2001]. In [Quellec et al., 201a-b] a different system of CBIR based on decision trees or an optimized wavelet transform provide other solutions.
Yet, our personal take is a different one in that we view CBIR and CBR as methodologies. Hence, *CBIR is a subset of CBR.* In this sense, our framework which combines the characteristics of CBIR and CBR into a hybrid reasoning approach is different than the one of [Jurisica et al., 2001]. We will still use the expression "CBIR-CBR combination", "hybrid CBIR-CBR" or the like, for the sake of reminding that we work with both.
For instance, the knowledge will not be structured by single element, but by cases, as it is done in CBR and as stated above, creating the knowledge case-base in an incremental manner. From CBIR's side we consider that image plays a prominent part in our approach, since medical assessment procedure can hardly work without it. From our point of view, it is also highly important to provide a continuity of the retrieval phase, starting with an efficient case adaptation and ending with the case storage and case-base maintenance. Both CBIR and CBR are context dependent. It is yet very difficult to overcome the context chasm, by propelling a generic solution, but some inner specific problem can be solved in a fusion paradigm. Hence the context modeling will be applied in our approach, related to our specific application. For instance, the manual procedure (as usually pathologists adopt) inconsistency will be alleviated by an automated grading provided by a semantic indexing method of histopathology images.
Our belief is that various query can be asked over a new case waiting for a solution. We consider that the retrieval results can be improved if we take queries into account and further more if we provide query expansion procedure over the new case. When it comes to selecting the most similar case, we propose a clustering of the obtain results to improve the accuracy of the retrieval phase in order to have an efficient case adaptation. In synthesis, the hybrid approach combines:

- **concept of similarity (central for both paradigms)**
- **visual content processing & analysis (CBIR)**

- **semantics, ontologies (CBIR)**
- **case-based structure of data (CBR)**
- **duality of medical knowledge (CBR)**
- **incremental learning (CBR)**

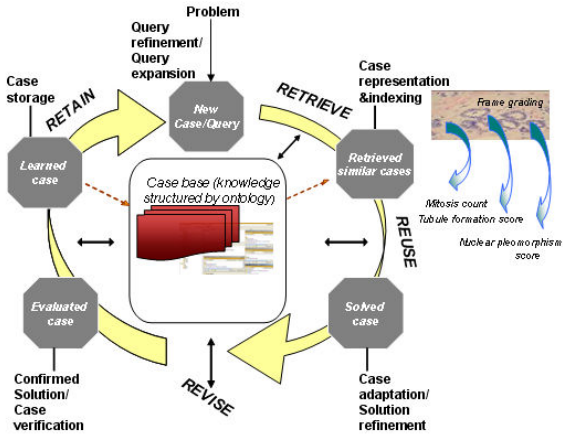


Figure 3.2. CBIR-CBR strategy

Figure 3.2 gives the insight of how the complete hybrid approach would function with all the phases implemented in the CBIR - CBR fashion [Tutac et al., 2009a]. This hybrid approach is proposed to be applied to the cognitive virtual microscope system which is discussed in chapter 8.

In order to answer the same question another direction we investigated was the semantic-dedicated representation approaches in medical domain.

Efforts to structure knowledge from different medical fields have been concentrated to define large vocabularies/ taxonomies, reference and application ontologies.
Physicians developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information efficiently. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge. Medical information systems, on the other hand, need to be able to communicate complex and detailed medical concepts (possibly expressed in different languages) unambiguously. This is obviously a difficult task and requires a profound analysis of the structure and the concepts of medical terminologies. But it can be achieved by constructing medical domain ontologies for representing medical terminology systems. Ontology-based applications have also been built in the field of Medical Natural Language Processing.

Semantic networks for instance, were applied to UMLS, SNOMED-CT [Bodenreider and Zhang, 2006], [Ouagne et al., 2005]. In the pathology domain, for lung disease

[Bontas et al., 2004] proposes formalised medical reports using UMLS concepts input represented in OWL language, in order to have a semantic-based retrieval for text and image data in a digital virtual microscope environment.
An evidence-based solution to assist in establishing a consensus in breast pathology domain is given by [Steichen et al., 2006], where SNOMED-CT terminology resource and available diagnostic classification system are used as a basis for building an ontology of morphological abnormalities (e.g. ductal carcinoma in situ). The NCI thesaurus is a logic-based domain description [Golbeck et al., 2003] and GALEN [Wang and Parsia, 2008] and FMA [Zhang et al., 2006] have been recently translated into OWL-DL to benefit of the computational advantages. In [Golbreich et al., 2005] the research goes further by proposing ontologies with rules for a brain anatomy ontology and investigates the reasoning support that is required.
From reference ontologies one could derive application or lightweight ontology thus being able to foster the integration, interoperability and reusability of shared knowledge. According to [Pisanelli, 2004], the representation and (re)organization of medical terminologies represents the main focus of ontology usage in medicine. He also gives a list of the most known medical ontologies designed and used today. A novel methodology and framework for structuring medical knowledge called Open Medical Development Framework (OMDF) was proposed in [Ouagne, 2009].

As one can see the need to structure knowledge in a formalized way, understandable by humans and computer agents, leads to ontologies in the breast cancer pathology as well [Dasmahapatra and O'Hara, 2006].
Thus in light of the arguments from above, CBIR-CBR combination (like we mentioned previously CBIR can be viewed as a subset of CBR, since they are both methodologies in our approach) is a novel direction for BCG, in which we decided for an ontological orientation.
In point of fact, ontologies can connect with CBIR. [Wang et al., 2006] proved that ontologies do help and improve the image retrieval process by narrowing the semantic gap which is without any doubt one of the key role the ontologies play in the semantic web. A study of precision results for Google search, an ontology text-based retrieval and multi-modality ontology-based retrieval (semantic and image information) is given in this work. The dataset consists of 4000 images from which the top 200 are evaluated (the application being oriented on canine subspecies). The experimental results show that the text-ontology performance is slightly better than the keyword-search, which is due to the lack of textual information in the web pages, while multi-modality ontology retrieval outperforms both keyword search and ontology-text retrieval. In most cases, ontology-based image retrieval can achieve better precision than keyword search, while in particular by combining high level semantic features with low-level image features, the precision finds an improvement by 5% to 30/%. Another work concerned with narrowing the semantic gap of CBIR using ontologies is proposed by [Iskandar et al., 2007] in which SPARQL queries on semantic representation of low-level image features provide a good retrieval result for 452 comic strips with tests on 1115 regions from 202 panels (each comic strip having one to five panels) . They also discuss the work of [Mezaris et al., 2003] which is also based on queries oven an ontology which describes high- level and low-level features of 5000 images from Corel library. The precision is shown to be better than recall, the overall results proving that an ontological approach performs significantly better, in addition to being more flexible than conventional text-based retrieval methods. This work is extended in [Mezaris et al., 2004].

In the same vein, another work proposed an ontology-based image annotation and retrieval approach [Styrman, 2005], while [Vacura et al., 2008] provides an alternative to MPEG7 standard, by developing a COMM ontology to describe low-level features of images.
Consequently, the new issue was how to connect ontologies with reasoning? Based on the survey given on formal representation and reasoning in chapter 2, we brought the pieces together and we realized that:

> *Ontology-type representation with reasoning features is the key to structured, meaningful and computable representation of knowledge from real world domain.*

This means that our purpose is to benefit of the *advantage to unify and structure* the knowledge from the medical real world-domain, providing *high expressivity* (OWL+SWRL) on one hand, and the capacity to *do reasoning based on logic formalism* (DL formalism), on the other hand.
Otherwise stated, we need *ontology reasoning* due to the following characteristics it has:

- helping design and maintenance of high quality ontologies
  - correctness (able to capture intuitions of domain experts)
  - richly axiomatized (sufficiently detailed descriptions)
  - minimum redundancy (no unintended synonyms)
  - meaningfulness (all defined classes can have instances)
- answering queries over ontology classes and instances
- integrating and aligning multiple ontologies

Large and generalized description frameworks provide a wide representation of a domain. However, it is often difficult to perform reasoning, particularly when a high level of computation is implied. Instead, application ontologies support this feature. Building multiple application ontologies with reasoning power and integrating them into reference ontologies could be a solution to this issue.

KR approaches	Advantages	Disadvantages	Breast pathology
non-logic-based formalism [Steichen, 2006]	human mind task-solving resemblance	lack of logical inference	semantic networks • large vocabularies (UMLS, SNOMED-CT) • reference ontologies (NCI thesaurus)
logic-based formalism [Baader et al., 2007]	high expressivity computational power	undecidability in complex representation	Description Logics (DL) • reference ontologies (GALEN)

Table 3.4. Knowledge representation approaches in breast pathology

Table 3.4 gives a synthesis of knowledge representation approaches in breast pathology, along with advantages and drawbacks, as this is the domain at which we focus further on.

3.3. Spatial Representation in Biomedical Ontologies

An important issue in the medical research and clinical practice is related to the role played by images: the investigation, the diagnosis and the prognosis, all are given based on the images- the samples taken at the patient's medical exam. They contain low level features and high level features, namely, concepts and relations between concepts. The spatial concepts and relations are often intrinsic to medical images. Therefore, a spatial representation and reasoning is crucial. We formulate this assumption as following:

In the medical field, spatial level requires spatial representation and reasoning.

When connecting the semantic level with the image low level, in a complex description, an extension of classic symbolic reasoning to visual reasoning is required.
One issue of great significance consists of defining spatial ontological categories when modeling ontologies. Such categories are parthood, location, connectedness, adjacency, etc. A roadmap for spatial representation has been proposed in [Bateman and Farrar, 2005], as part of the complex OntoSpace project. Based on the principle that space is an ontological category, it analyzes five different projects on the ground of the mereological, topological and geometrical descriptors. It also discusses the notion of scale with respect to granularity in the BFO (Basic Formal Ontology) space [Grenon and Smith, 2004], particularly in the SPAN spatial-temporal subtype of it.
Yet the spatial information encapsulated within a biomedical ontology is often ambiguous and leads to inconsistency in the process of reasoning, even driving to a situation where no model could be found.
We postulate that:

Biomedical ontologies with spatial representation need formal spatial theory support.

To this end, a comprehensive approach to introduce formal theory for spatial representation and reasoning in biomedical ontology is presented by [Donnelli et al., 2005]. It tackles mereo-topology spatial concepts, (as discussed in chapter 2, mereo-topology is a fist-order theory which embodies qualitative mereological and topological concepts (e.g. parthood, location).
However, the paper focuses only on mereology -basic parthood and location relations in a detailed analysis of FMA and GALEN biomedical reference ontologies.

A complementary contribution for how to distinguish between parthood and location is given by [Schulz et al., 2005]. An example of application ontology of radiological anatomy is provided by FMA -Radlex [Mejino et al., 2008], where the limits given by the ambiguity in usage of location relationships *part-of , contained-in, is-a* from RadLex alone are alleviated through FMA derivation.
The spatial reasoning is not fully automated, yet obviously inextricably intertwined with the representation itself in both the roadmap and the formal theory mentioned above.
An example of a spatial reasoning with OWL-DL is given in [Mechouche et al., 2009], for a semantic annotation of gyri and sulci parts of the brain MRI images. It combines symbolic knowledge (ontology) with numerical atlas knowledge and annotation is based on mereo-topological relations.
Another way to represent spatial information is introduced by [Hudelot et al., 2006] which emphasizes on topological and metrical relations, whereas [Mezaris et al., 2004] focuses on geometrical descriptors, in which the link between region and semantic description of objects is addressed by a low-level to intermediate-level descriptor mapping.
Table 3.5 presents a compilation of the spatial approaches in biomedical ontologies, emphasizing their advantages and shortcomings.

Spatial approaches	**Biomedical ontologies**	**Advantages**	**Drawbacks**
mereology [Donnelli et al., 2005], [Bateman and Farrar, 2005] [Schulz et al., 2005] [Mechouche et al., 2009]	▪ FMA ▪ SNOMED-CT ▪ GALEN	▪ reduces ambiguities ▪ symbolic knowledge &numerical Knowledge ▪ link with image level	▪ decidability issues at reasoning (large vocabularies)
topology [Hudelot et al., 2006] geometry [Mezaris et al., 2004]	▪ FMA (brain MRI images) ▪ general purpose	▪ image interpretation	▪ reasoning capabilities

Table 3.5. Spatial approaches in biomedical ontologies

3.4. Conclusions

This chapter aimed at firstly presenting the state-of-art literature in knowledge representation and reasoning in medical applications. We discussed the approaches for knowledge representation and reasoning starting from CBIR and CBR which paved the way for the introduction of biomedical ontologies.

Lastly, we focused on visual representation and reasoning in ontologies from biomedical field. We showed the red thread based on which we built our assumptions.
Given this basis, we can conclude that merging OWL with DL formalism for having a reasoning support on the knowledge representation, especially applied on the biomedical ontologies is a very efficient way of doing semantic web and thus narrowing the semantic gap. The key idea is to keep the balance between computational power and high expressiveness. This becomes very important when to integrate SWRL rules with ontologies, as the limitation is given by the DL reasoner capabilities.
Nonetheless, ontologies are capturing more and more interest in ubiquitous domains, as it is considered the latest trend in structuring and formalizing knowledge.

One of the medical domains of high significance nowadays is the domain of breast cancer grading or the histopathological grading. We introduce this domain in the following chapter. What is to be noted is the fact that despite its relevancy there is not a single ontological representation of breast cancer grading and there is no spatial representation and reasoning approach for this domain.

Hence, our objective is to advance a novel ontological model for breast cancer grading with reasoning and rule module. Furthermore, we also propose a formal basis for the spatial representation of the histological grading.

A remark worth mentioning in closing the chapter: if we are to look at ontologies from a histopathological perspective, ontologies make the *nucleus*, the *organelle* of Semantic web, which are constituent elements of the future Web *cells* generation [Damjanović et al., 2003]. To avoid having a *malignant tissue* of the *new Web organisms*, an appropriate modeling and reasoning of these *semantic cells* is required.
In the end of this chapter, we summarize the contributions and research directions that arouse from the study presented in this chapter:

- comparative analysis of CBIR and CBR in medical applications
- ontologies as a solution to narrow the semantic gap met in CBIR [Iskandrar et al., 2007], [Wang et al., 2006] [Mezaris et al., 2004], [Mezaris et al., 2003] and to formally represent knowledge in a highly expressive way with DL reasoning powers (using OWL and SWRL languages) [Baader et al., 2007], [Mejino et al., 2008]
- need for spatial theory support in biomedical ontologies based on the study of spatial representation and reasoning in medical applications [Mechouche et al., 2009], [Hudelot et al., 2006], [Donnelli et al., 2005], etc.

4. A Formal Representation Model for Breast Cancer Grading

As mentioned in the previous chapter, a formal semantic representation is required in order to have a structured and consistent knowledge to be able to perform medical assessment. We assign this chapter to present the novel semantic approach for a particular medical application, the breast cancer grading, namely. We discuss the issues that the breast cancer grading has and we show how through ontological representation these problems are overcome. The ground on which we model the representation is set up in section 4.2, followed by the presentation of our methodology.

4.1. Breast Cancer Grading

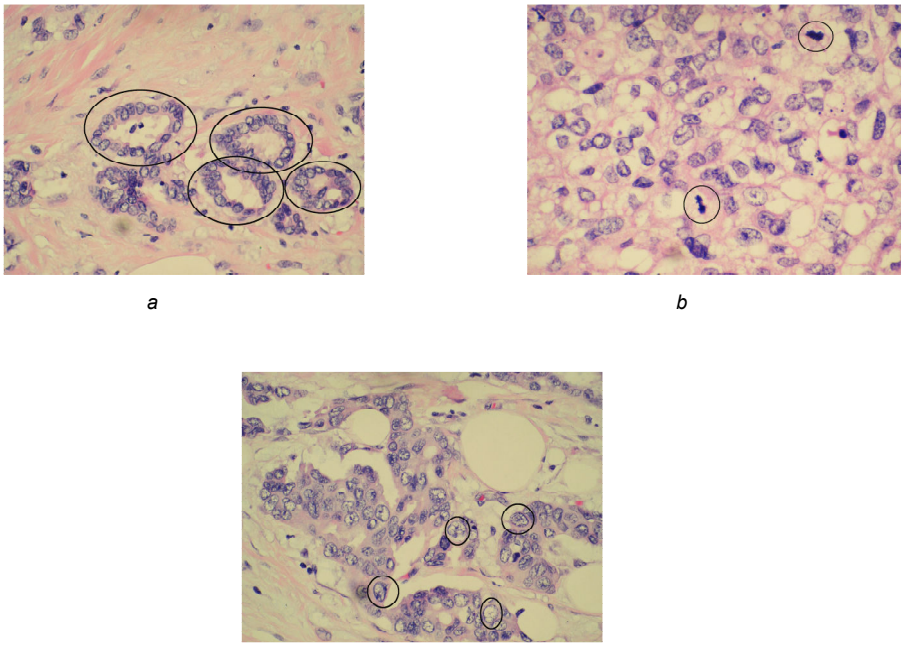
Breast cancer refers to a malignant tumor that has developed from cells within the breast. The breast is composed of two main types of tissues: glandular tissues and stroma (supporting) tissues. Glandular tissues house the milk-producing glands (lobules) and the ducts (the milk passages) while stroma tissues include fatty and fibrous connective tissues of the breast. The breast is also made up of lymphatic tissue-immune system tissue that removes cellular fluids and waste. Breast cancer is a leading cause of death among women, and its incidence is rising. Although curable, especially when detected at early stages, breast cancer is expected to account for 28% of incident cancer and 20% of cancer deaths in women. There are several types of tumors that may develop within different areas of the breast. Most tumors are the result of benign (non-cancerous) changes within the breast.

Histological grading is nowadays considered an assessment of high relevance in breast cancer prognosis of modern pathology. It is a microscopic image-based prognosis as it deals with histopathologic images analyzed under a microscope. Most grading systems currently employed for breast cancer combine criteria in nuclear pleomorphism, tubule formation and mitotic counts. In general, each of three elements is assigned a score on a scale of 1 to 3 and the final grade is determined by the sums of the scores [Cardiff and Jensen, 2000].

Until recently, the most common grading systems used in the United States were the original Scarff-Bloom-Richardson (SBR) system and the Black method which emphasizes nuclear grading and excludes consideration of tubules as criteria. In Europe, the Elston-Ellis modification of the SBR grading system (Nottingham Grading System) is preferred and is becoming increasingly popular in the US. This modification provides somewhat more objective criteria for the three component elements of grading and specifically addresses mitosis counting in a more rigorous fashion. For example, hyper chromatic nuclei and apoptotic cells which are counted in the original SBR system are excluded in the NGS and the area being assessed is specifically defined in square millimeters. These modifications have enhanced reproducibility of grading among pathologists and to a considerable extent have fostered acceptance of grading by clinicians.

To synthesize, among the standard grading systems, Nottingham Grading System (NGS) represents the gold standard (ground-truth) due to its objectiveness for the three components of grading. The three components of NGS criteria are briefly described below (see Figure 4.1, Table 4.1):

- Tubule Formation score (TF) - are referred as the density of the Tubule Formations - white blobs (lumina) surrounded by a continuous string of cell nuclei
- Mitosis Count (MC) score represent the number of Mitoses - diving cells nuclei. MC is assessed in the peripheral areas of the neoplasm and it's based on the number of mitoses per 10 High Power Field's (HPFs) – high resolution (usually 40X) frames obtained using microscopic acquisition.
- Nuclear Pleomorphism Score (NPS) - categorizes cells nuclei based on two main features: size and shape. If a histological frame contains a mixture of nuclei population with different grades, the highest score is taken into consideration.



a *b* *c*

Figure 4.1. NGS components : a) Tubule formation: lumina surrounded by string of cell nuclei b) Mitosis: dividing cell nuclei c) Big size/irregular shape nuclei-NPS grade 3

BREAST CANCER GRADING with NGS		SCORE
NUCLEAR PLEOMORPHISM (NP)	Small Regular Uniform Cells	1
	Moderate Nuclear Size And Variation	2
	Marked Nuclear Variation	3
TUBULE FORMATION (TF)	Majority of Tumor (>75%)	1
	Moderate Degree (10-75%)	2
	Little or None (<10%)	3
MITOTIC COUNT (MC)	0-9 Mitoses/10 hpf	1
	10-19 Mitoses/10 hpf	2
	20 or > Mitoses/10 hpf	3
COMBINED HISTOLOGIC GRADE	Low Grade (I)	3-5
	Intermediate Grade (II)	6-7
	High Grade (III)	8-9

Table 4.1. Nottingham Grading System

Breast Cancer Grading requires time and attention, dealing with hundreds of cases by day, each of them having around 4000 digital frames at 40X magnification or exploring them at low level magnification. Currently, BCG is achieved by visual examinations of pathologists. Such a manual work is time-consuming and inconsistent, according even to the pathologists' opinion. The grading performed on the same slide of the same patient by different pathologists (inter-observer agreement) or at different time by the same pathologist (intra-observer agreement) can differ. These agreements provide the data to compute the kappa-coefficient (k-coefficient).

These inconsistencies are mainly related to the subjective manual scanning and evaluation of the mitosis, tubule formations and the cells nuclei, the experience and stamina of each medical expert. Considering these drawbacks, different solutions have been proposed. Most of them addressed the direction of CAD (Computer-Aided Diagnosis) by developing a semi-automatic grading system. The idea is to assist in the process of grading with an automatic or semi-automated algorithm with the help of a computer. To this end, several approaches have been developed considering only individual parts of the BCG. Automated nuclear pleomorphism score was proposed by [Demir and Yener, 2005], [Jeong et al., 2005], [Adawi et al., 2006], while tubule formation score was addressed by [Petushi et al., 2006] and mitosis count by [Beliën et al., 1997].

Yet, no attempt has been done to combine all criteria in order to provide a complete automated BCG and further more no attempt has been done to combine the image processing algorithms with the semantic level. Therefore, we propose a solution to meet pathologist's needs with a novel BCG, thus alleviating the shortcomings of the manual grading procedure and further more overcoming the context gap and the semantic gap discussed by [Smeulders et al., 2000].

4.2. Problem Formulation

Formal representation embodies different ways of mirroring facts from real world domain. Two major aspects need to be discussed here.
Firstly, ***the nature of formal representation: qualitative representation versus quantitative representation***. In chapter 2 we addressed the advantages and drawbacks of both representations. We emphasized the idea that *quantitative approaches sometimes force the use of quantities to express even qualitative facts*. On the other hand, the very conception of OWL formalism leads to the purpose of the language itself and that is to support *logical qualitative definition of classes and not quantitative definitions.*
Coming to our application domain, it is obvious that the NGS contains a combination of both types of information: quantitative information, such as scoring or grading and qualitative such as „dividing cell nuclei not located in tubule formation area, close to neoplasm periphery". Hence, the challenge is reconcile these two approaches. To this end, we propose the following tenet based on which we build our model:

Reflection of the breast cancer grading domain is carried by means of qualitative formal representation without resorting to the traditional quantitative techniques. Where met in the breast cancer grading domain, quantitative definitions are confined to the qualitative representation as numerical values allowed by semantic languages.

Secondly, ***the target of representation and the ways of representation***. In view of the reasoning characteristics and challenges, our purpose is to create an ontological model for breast cancer grading that will address the following issues:

1. Endurants representation (objects from the breast cancer domain such as microscopic spatial objects, lumina, tubule, mitosis, nuclei, DCIS but also other objects seen as concepts). We do not model perdurants (events, processes). This consideration entails the assumption of a time-independent representation, as we only focus on a spatial extension and not on temporal. The reason is that we regard time in terms of morphogenesis (e.g. the mitotic phase - the process of cell nuclei division) rather than strictly a matter of breast cancer grading assessment.
2. Endurants are symbolically represented using OWL-DL and SWRL formalism to achieve high expressivity. This constitutes a clear illustration of qualitative formal representation in which OWL as well as SWRL constructs contribute to the reflection of the breast cancer grading domain. Numerical values are given in a qualitative framework.
3. However, working with SWRL rules leads to undecidability, and hence there is a tradeoff between expressivity power and computational power. To overcome this problem, SWRL DL safe rules are introduced.
4. Formal spatial theory support for spatial relationship encountered in breast cancer grading in order to eliminate ambiguity and inconsistencies in

representation. The satisfiability of concepts by tableau-based algorithm (finding clash-free models) is verified for spatial representation and the entire semantic description.

4.3. Methodology for Ontology Modeling

We coin the term *semiologic approach* for the breast cancer grading, considering the fact that it deals with the symptoms of the disease, how abnormal the cancer cells from the tissue appear under the microscope. It answers to the question: *what* can be seen, the objects from within the histopathology images. Hence, our goal is a semiologic formal representation of perdurants, a breast cancer grading ontology (BCGO.) In this way, the grading of breast carcinoma could be integrated into upper-ontologies of breast pathology. We envision this breast cancer grading ontology would be very helpful when performing the prognosis assessment. It complements the standard grading system (NGS) with terminological consistent description, serving as a virtual second opinion in grading, which is important for the pathologist's reliability and to increase confidence in the diagnosis. Since the medical communities are open to standardize their terminologies for general purposes but also for dedicated applications and they view ontologies appropriate for this purpose (UMLS, SNOMED-CT, FMA, GALEN, NCI/NIH), we consider that a breast cancer grading ontology would follow along this line. Additionally, the use of automated processes in various areas of medicine for diagnosis, and the directions proposed by the Virtual Physiological Human (VPH) community (see section 9.2), we envisage that the medical community will be more opened in the future to use ontology-driven second opinions in their work. Clinical cases organized under BCGO terminology allow the pathologists to find similar cases when searching in the knowledge-base in order to perform grading.

There are various ways of designing ontologies [Gruber, 1995]. The study of [Lopez and Perez, 2002] gives an exhaustive overview of methodologies for building ontologies. However, no single right methodology has been acknowledged. Therefore, we adjust the principles of Uschold and Grüninger [Uschold and Grüninger, 1996] to fit our BCG specific application. We follow two paradigms to associate meaning to features extracted from the image, thus indexing images by semantic means. We call the first, the image-driven approch and the second, the concept-driven approach.

In the image-driven approach, the emphasis is on the semi-automated breast cancer grading, thus the idea is grappling mainly with the context gap, as shown in Figure 4.2.

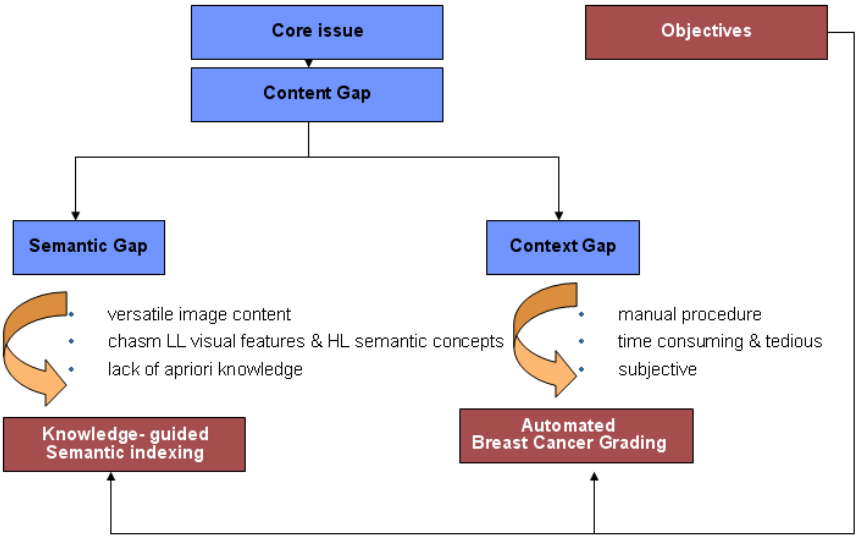


Figure 4.2. Our solution to tackle with the content gap

Our paradigm workflow depicted by Figure 4.3 follows the next steps [Tutac et al., 2008b]. The first step consists of digitizing the histopathology slides analyzed under the microscope by the pathologist. As mentioned previously, the slide consists of around 4000 frames which will be processed and analyzed in the next step, by integrating the medical knowledge. To have a complete domain knowledge analysis, *duality of objective knowledge and subjective knowledge* is required. The subjective knowledge (coming from the pathologist) as well as the objective knowledge (coming from the Nottingham Standard Grading System) are structured in a formal representation within our BCGO.
Medical knowledge concepts and rules are then translated into computer vision (CV) concepts and rules based on the Generic Translator Framework (GTF), both required for the image processing and analysis step .The output represents the semi-automated Breast Cancer Grading which can be used in the diagnosis and prognosis assistance.

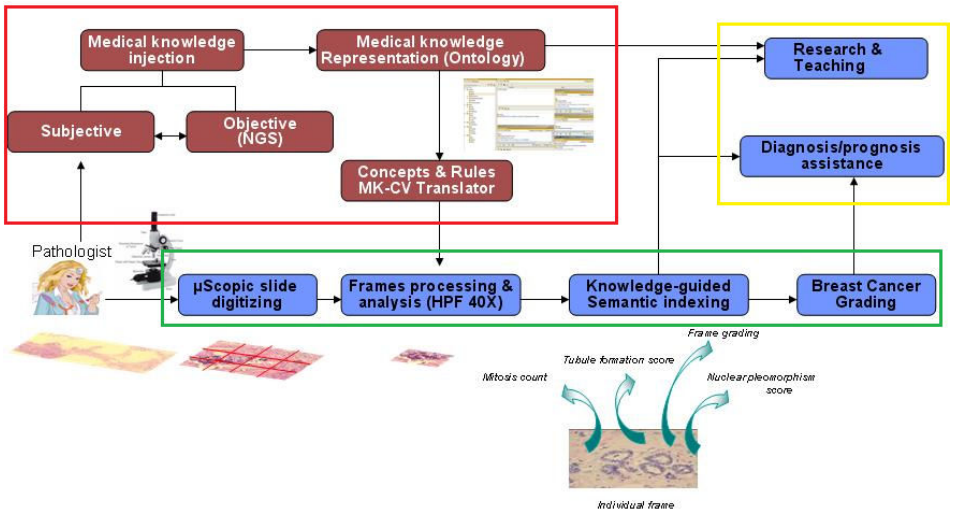


Figure 4.3. Knowledge-guided semantic indexing workflow

As shown by Figure 4.4, the GTF is structured in three parts: development of the correspondence between medical concepts and computer vision concepts, definition of intermediate CV rules and generation of the final Symbolic Rules by fusion of the CV concepts and intermediate CV symbolic rules [Tutac et al., 2008b]. Note that CV concepts are used as an input for the Rules Translator together with the medical rules to generate the CV preliminary symbolic rules.

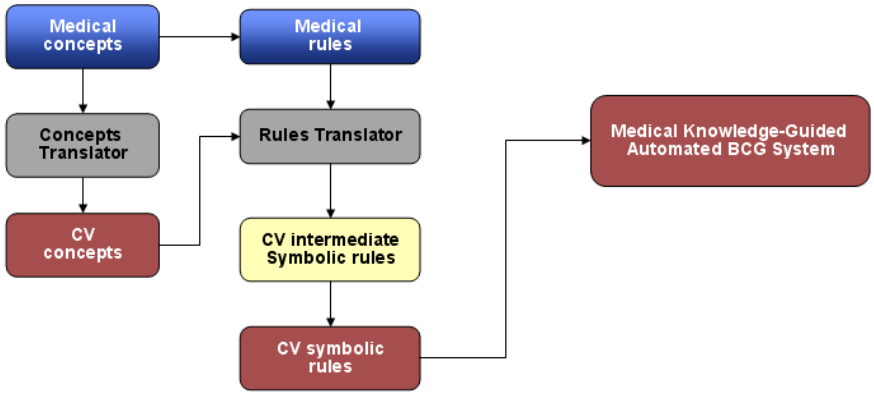


Figure 4.4. Generic Translation Framework

The MK-CV concept translator is based on a classification of elements that need to be taken into consideration to give the final grading.

- objects : Image, Cells, CellsCluster, Lumina, Tubule, Mitosis, Nuclei Pleomorphism;
- attributes: size, shape, intensity, localization;
- values : small, medium, large, regular, variated, irregular, very high, dark, very dark, eccentricity;
- operators: the quantifiers provided by the DL syntax

An illustration of objects translation is given by Table 4.2 [Tutac et al., 2008a], [Tutac et al., 2008b].

Medical Objects	CV objects
Slide	Image (digitized)
Cells	Cells
CellsCluster	Union of Cells
DarkCellsCluster/ VeryDarkCellsCluster	Union of Cells (with color feature segmentation)
Lumina	White compact segments of the Image included in the union of dark cells
TF/Mitosis/NP	Union of Cells/Diving Cells nuclei/ dimension & shape features of the nuclei

Table 4.2. MK-CV objects (of concepts) translator

To better understand the MK-CV rules translation process, an illustrative example is given in the case of the mitosis definition.

- Mitosis Pathologist Rule :

  very dark diving cells nuclei from the peripheral area of neoplasm

  ↓

- Mitosis CV intermediate symbolic Rule :

  ↓ $$VeryDarkCells = \{\bigcup Cells_j \mid intensity(Cells_j) \leq VeryDark\} = \{VeryDarkCell_j\}_{j=\overline{1,J}}$$

- Mitosis CV Symbolic Rule :

$$M_{ROI} = \{\bigcup VeryDarkCell_j \mid ecc(VeryDarkCell_j), size(VeryDarkCell_j), VeryDarkCell_j \not\subset TF_{ROI}\}$$

Figure 4.5. MK-CV rule translator

The problem with this approach is that we could not fully benefit of the DL formalism's advantages as it is too close to the image processing level. Although there is, in a sense, a knowledge-guided semantic indexing, (as shown in Figure 4.3), the ontology in fact does not play the pivotal role. The connection between medical knowledge and image processing and analysis is much stronger. This,

moreover, creates difficulties for a medical expert to understand it and further give any refining advices. Based on these rationales, we take the second approach.

In the concept-driven approach, we focus on the role of formal representation in order to capture the knowledge as close to the medical language as possible. The semantic representation will guide the image processing level in that sense rather than being two independent steps.
This time in the process of modeling the BCGO we identify a three-phase procedure [Roux et al., 2009a]: knowledge acquisition, knowledge translation and knowledge refining, as shown in Figure 4.6. A top-down development process starting with the definition of the most general concepts in the domain is subsequently followed by specialization of the concepts and addition of properties.

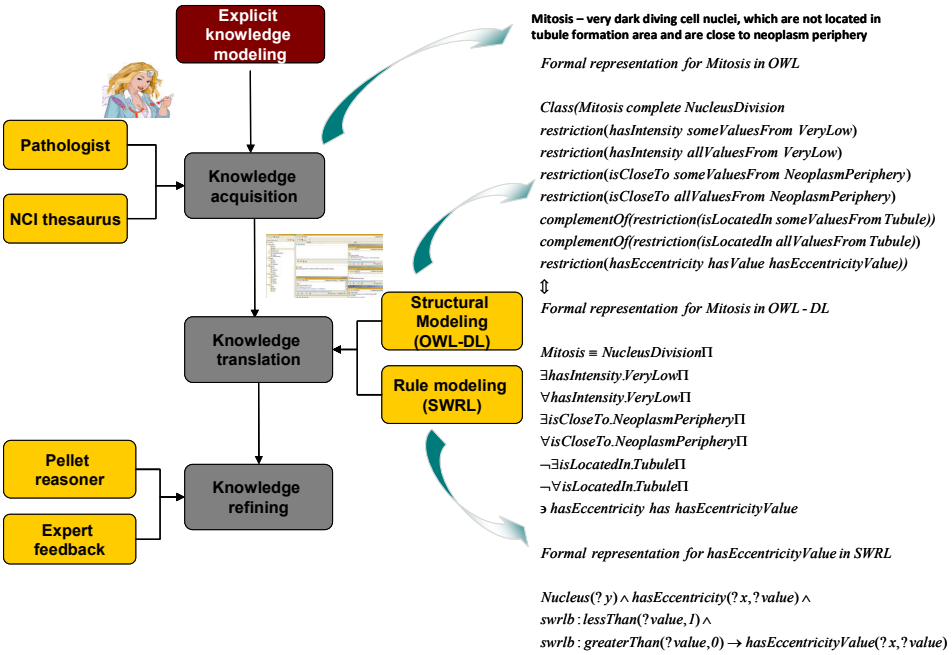


Figure 4.6. BCGO knowledge-modeling methodology

Based on the strong points the second approach has, we hold to it and we further present the three step procedure in detail.

4.3.1. Knowledge Acquisition

As an important prerequisite, we define the domain of representation as Breast Cancer Grading (BCG) and we plan to use this ontology to overcome the subjectivity issue of decision making in the grading.

The knowledge acquisition has two inputs: the NCI (National Cancer Institute) thesaurus and the pathologist.

The reasons we adopt NCI thesaurus and not SNOMED-CT, UMLS or Medical Subject Heading (MeSH[2]) standards are multiple:

- Homogeneity of knowledge: SNOMED-CT, UMLS contain heterogeneous information, while NCI thesaurus is specialized on describing cancer concepts.
- Information level: NCI provides formalized knowledge for cellular levels which is relevant to the breast cancer grading.
- Inter-operability factor: semantic alignment, semantic mapping (our purpose is to link the breast cancer formalized knowledge from the NCI thesaurus with breast cancer grading knowledge).
- Logic-based factor: NCI thesaurus is a logic-based domain description (as we mention bellow). This is a considerable advantage for our approach is which DL plays an important role.
- Usability factor: NCI thesaurus is freely available, while SNOMED-CT for instance is not.

For these reasons, we did not delve into a more detailed comparison between UMLS, SNOMED-CT, MesH and NCI thesaurus. However, the NCI thesaurus is also a very large and complex ontology and we do not need all the concepts within it in our ontology. Therefore, a process of segmentation is needed. In our first approach, we performed a semi-automated segmentation using PROMPT[3] plug-in of Protégé framework (the framework we use to formalize our ontology) and the Pellet DL reasoner to classify the concepts. The fact that NCI thesaurus has been translated into OWL-DL format [Golbeck et al., 2003] and [Foy et al., 2007] helps us benefit of the DL reasoner capabilities. The procedure consists of the following steps:

- Manual selection of which concepts to include in the sub-ontology using *medical criteria*, not computer science criteria (e.g. bacteria may be excluded from the sub-ontology because by virtue of its *medical definition*, it is not directly relevant to breast cancer). An identification of irrelevant nodes from within a relevant breast cancer concept is also manually performed.
- PROMPT automated process extraction
    - Each concept is treated separately. After the segmentation is accomplished the results (which are already in OWL-DL format) are classified and checked for consistency by the reasoner.
    - Extraction of the matching super-classes of the deepest relevant nodes until the subset is complete.
    - The process of segmentation is repeated for all relevant concepts which are eventually integrated into our breast cancer grading ontology
    - Medical and OBO validation (which is detailed in chapter 7)

In the approach of [Seidenberg and Rector, 2006], several automated methods are proposed to extract and evaluate the segmentation of relevant fragments out of large description logic GALEN ontology. Our next approach will use these methods applied to NCI ontology. In section 8.3 we also discuss about the segmentation of objects related to BCGO concepts such as nuclei, tubule and mitosis.

[2] http://www.nlm.nih.gov/mesh/, last accessed October 2010
[3] http://protege.stanford.edu/plugins/prompt/prompt.html, last accessed July 2010

Hence, the subset of concepts from the NCI thesaurus used in our ontology contains generic concepts such as *Disease*, *Patient* and specialized concepts like *DuctalCarcinomaInSitu*, *CancerPatient,* etc.

Nevertheless, NCI does not handle specific concepts for grading, therefore to make the knowledge acquisition complete, we rely on the pathologists to obtain the information, the medical cases and to assist the process of grading. Since the pathologists use the Nottingham Grading System as standard we adopt it as well in the knowledge representation and we add concepts such as *Assessment, TubuleFormationScoring, Mitosis, Tubule, NuclearPleomorphism, etc.*

4.3.2. Knowledge Translation

In our view, the term *knowledge translation* essentially refers to translating the medical information related to breast cancer grading into a formal ontology-like representation. Different from our first perspective [Tutac et al., 2008a], [Tutac et al., 2008b], this phase unfolds into two steps OWL-DL for structural modeling and SWRL for rule modeling. The reason we rely on the two modules is that our purpose is to achieve high-expressivity and decidability in the same time. In this way, we practically define a hybrid ontology. These modules can be used either separately or together in describing classes and properties of the ontology.

Structural modeling refers to defining the structure of the knowledge, in terms of its constituents and the type of connections between them. As stated in [Guizzardi, 2005], the structural modeling of conceptual models works with *endurants (objects) and perdurants (events, processes*).
Our ontology addresses and models the endurants, since for the events and processes a temporal extension would be required. Nevertheless this is a matter of future interest for us, as one very illustrative example from the breast cancer grading standard system is the one of mitosis criteria, which is by definition a nucleus division, a division that happens at a certain moment of time.

Rule modeling. As our objective is to gaining high expressivity, we resort to rule modeling at that point where the OWL-DL lacks the power of expressing more specific and detailed information. This is the junction where we need an adaptive and interactive approach – thus, we connect to SWRL. This behavior has a major advantage: it is not a black-box type of building the rule model. It is transparent and with the help of reasoning engines, one can check whether the rule has been edited correctly from the syntactical point of view and after that, processed. Furthermore, a possibility to check the semantic correctness of the information – introduced as rule – is to use queries. By retrieving query answers, it is possible to check the interpretation of the concepts already stored in the ontology.
On the other side, this kind of hybrid modeling is more complex due to the decidability issues. As mentioned previously, it is generally acknowledged a combination OWL-DL with SWRL rules is usually undecidable and therefore, the reasoning support is of high importance. However, the Pellet reasoner [Sirin et al., 2005] is able to handle DL safe rules and the reasoning algorithm can finish in a finite time (it provides a model as output).
With regard to connecting the semantic level with the image level, we rely on the correspondence between ontology language and programming languages (image

processing) given in Table 4.3. This correspondence is more general and rigorous than the MK-CV translator from the image-driven approach.

Ontology Language	**Programming languages**
class	class
property, attribute	variable
instance, individual	object
value of property, attribute	value of variable
domain of property, attribute	class that contains the variable

Table 4.3. Ontology Language versus Programming Languages

4.3.3. Knowledge Refining

By knowledge refining we mean two things: ontology revision from both formal representation and clinical perspective. This revision implies reasoning. During the modeling of the ontology, satisfiability and subsumption checks regarding concepts, as well as *AB*ox and *TB*ox consistency tests are performed via the Pellet reasoning service. The automated classification of class hierarchy (taxonomy) and the entailments checking both satisfy the coherence condition.
The reasoning makes use of the explicit knowledge represented in the ontology and makes the connections between classes, or properties, in terms of inference process. The reasoner constructs the asserted and inferred models which could be equivalent, if the reasoning could not infer any other statement other than the ones already explicitly given. Otherwise, the models differ.
Based on [Baumeister and Seipel, 2005] and [Corcho et al., 2004], Table 4.4 synthesizes the most probable issues encountered in ontology building, which request a refining of the ontology. We will discuss these issues in detail in chapter 6. After we completed the first step, an intervention from the medical side is also required when axioms need to be refined or completed, or rules need to be added. For instance, the restrictions related to the location of mitosis – at the peripheral area of the neoplasm – could be omitted from the description and thus the reasoning combined with the image processing could obtain objects different from mitosis as output; it may be nuclei with bizarre shape, yet not mitosis.

Inconsistency	Circularity issue	Observations
	Partition errors	common classes in disjoint decomposition& partitions
		common instances in disjoint decomposition & partitions
		external classes in exhaustive decomposition & partitions
		external instances in exhaustive decomposition & partitions
	Semantic errors	
Incompleteness	Concept classification	
	Partition errors	disjoint knowledge omission
		exhaustive knowledge omission
Redundancy	Grammatical issue	subclass of relations
		instances of relations
	Formal definition issue	identical for some classes
		identical for some instances

Table 4.4.Ontology refining issues

Therefore, adding this information in the mitosis definition practically means that mitosis is not located in the tubule formation area.

4.4. Conclusions

The aim of the chapter was to present a novel approach to narrowing the semantic gap and the context gap in the breast cancer grading domain. The rationale that stood at the foundation of this approach is both of medical and scientific perspective. The breast cancer grading along with its characteristics and problems was introduced in section 4.1. In the following section, we formulated the assumption and considerations based on which the formal model for BCG is built. Section 4.3 we draw together two approaches and we argue the limitation of the first approach that was taken in the previous stages of our research and that lead us to focus on the latter one.

The novelty of our approach consists of a qualitative representation in OWL-DL coupled with SWRL module formalism. As the breast cancer grading implies image analysis, spatial concepts and relations, another novelty we bring is a formal theory support for the spatial knowledge representation. The spatial representation is also carried in a qualitative manner.

Even though the ontology is designed to be an application-oriented ontology and not reference ontology, the methodology we adopt is constraint free, generalized and applicable to any other domain representation. Manifold applications arise from building such ontology.

The first that comes to mind remains in the domain of the semantic web and that is the integration in upper ontologies, fostering the mapping of concepts and the enriching of the domain with semantic representation on grading. Another application regards the integration of the ontology in a cognitive virtual microscope platform, which is however one of our objectives. The integration contributes to providing reliable assistance to the pathologists in the grading assessment.

In conclusions the contributions we bring in this chapter are summarized as following:

- generalized methodology for ontology building consisting of three step procedure : knowledge acquisition, knowledge translation and knowledge refining, respectively
- Breast cancer grading application ontology BCGO instead of reference ontology (as a semiologic model)
- qualitative representation approach instead of quantitative representation (based on the arguments given in chapter 2)
- perdurants modeling (objects not processes or events)
- spatial representation extension for the BCGO model
- analysis of two approaches for the BCGO model: image-driven approach and concept-driven approach. The analysis evidenced that the latter fits our purpose objectively, therefore we further apply it.

5. Semantic Reasoning for Breast Cancer Grading Model

In the context of semantics and imaging, one major concern deals with the relation between semantic reasoning and spatial reasoning. This is important due to the fact that the some of the semantic concepts are described by means of spatial concepts.
In this chapter we propose a novel approach in which spatial reasoning is integrated into semantic reasoning with the formal representation of the breast cancer grading domain knowledge. In other words, the spatial representation formal theory could be seen as applied to the process of ontology design.
We firstly set a theorical foundation to formally describe the spatial knowledge and we further present how to perform reasoning based on it.
Representing spatial concepts and relations based on a formal theory allows us to eliminitate ambiguities and inconsistencies in the formal representation and reasoning with the semantic concepts and relations.

5.1. Formal Theory for Spatial Representation

Our general approach follows the steps shown in Figure 5.1. The spatial concepts are the input for formulating the spatial claims which are verified based on the spatial axioms and theorems in the spatial reasoning phase, either in the manual or automated way. The reasoning provides the inference results which will show if the representation is either consistent or inconsistent.

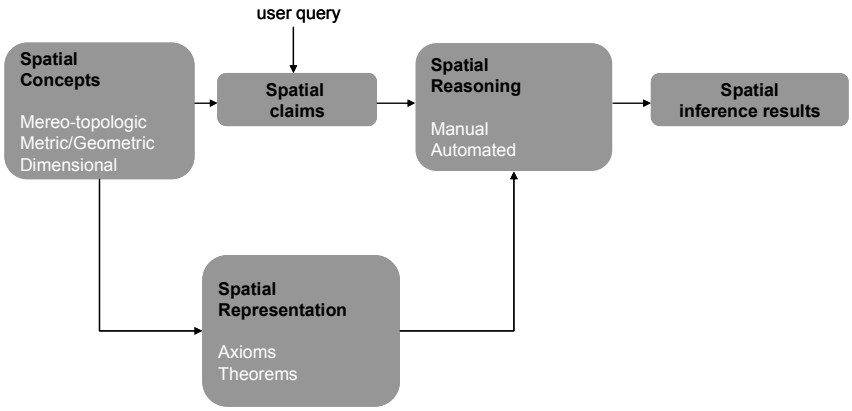


Figure 5.1. Spatial representation and reasoning approach

Before introducing the formal theory, we want to emphasize several important aspects based on which we build our approach.

Definition of spatial concepts nature. There are different perspectives on defining spatial concepts. In some approaches, they are viewed as concepts by their own such as in [Hudelot et al., 2008], in others, they are viewed specifically as properties [Brageul and Guesgen, 2007]. Lastly in other cases, they are used interchangebly [Donnelli et al., 2005].
In the first situation, the concepts would correspond to classes according to OWL terminology and therefore a definition would be required, possibly including restrictions composed of properties.
In the second case, no other definition would be required, they are simply taken as properties and used in the description of clases or individuals, or as relations between classes/individuals, according to the knowledge given by the domain.
Our approach is different in the sense that, spatial concepts are viewed as concepts and they have properties associated with them (for instance, Surroundness concept and *isSurroundedBy* is the associated property which has Surroundness as range).

Type of spatial concepts. All ontological categories of things or processes of a domain are defined by their *qualitative features* and inter-relationships, not by *physical quantities* [Bateman and Farrar, 2005] including distance and orientation. A concept/class has some instances, which can only take a predetermined number of values. The inference rules use these values and not numerical quantities approximating them. Based on this rationale and considering that space is an ontological category, all spatial relationships we describe are qualitative.

Time as ontological category. As mentioned in chapter 4, we view time in terms of morphogenesis (e.g. the mitotic phase - the process of cell nuclei division) rather than strictly of breast cancer grading assessment. Therefore we apply the time-independent formal theory with the mereo-topological relations introduced by [Donnelli et al., 2005] to the domain of breast cancer grading. We further extend the formal theory to geometric and metric spatial relations, which are relevant to breast cancer grading.

Axioms of mereo-topology. Concerning mereo-topological relations, it is necessary to address the issue of axioms. In mereo-topological theory the relations are viewed as primitives hence the axioms are embedded within the theory. Every spatial concept must conform to the axioms. The most generally used axioms are reflexivity, transitivity and symmetry/antisymmetry. Hence, in regard to mereo-topological relations we will only discuss these kinds of axioms. For the metric relations we hold the same axioms of the mathematics.

Convention in notation. The convention we use for axioms and theorems is as following: type of relation we are referring to, followed by A for axioms or T for theorems (e.g. PA denotes the parthood axiom; ST1 denotes the first surroundness theorem). Additionally, we use x, y for pointing to individuals and A, B to classes. We make use of the OWL-DL syntax and conventions of representation. Also, in order to understand the definition and formulas for axioms and theorems, we make use of the symbols found in Table 2.7.

5.1.1. Generic Definitions

We begin with what we call, *generic definition*: the definition of the *SpatialObject* and the *SpatialRelation* concepts along with its derivatives, extending [Hudelot et al., 2008]. They are followed by the mere-topological, metric and dimension axioms and theorems.
The objects related to breast cancer grading anchored into image space are the anatomical entities (*AnatomicEntity*) and microscopic entities (*MicroscopicEntity*). We also consider the slide and frame concepts as part of the image space (viewed as *ConceptEntity* in our BCGO, as subclasses of *Virtual Specimen*).

Definition 5.1. Let *R* be the *SpatialRelation* - a binary relation between two classes A and B, where *R* is the combination of *r* relation (between individuals x and y) with *Instantiation* relation from a domain D.
The instantiation relation *Instance* holds between an individual x and a class A, if the individual x is an instance of A. The axioms of instantiation relation are given in *InstA2* and *InstA3* according to [Donnelli et al., 2005]. Note that *Instantiation* relation is not a spatial relation, but is needed to reflect all subsumption relations (the instance checking is important in the reasoning phase, on the tasks related to *AB*oxes).

$$
\begin{aligned}
&(InstA1)\ Instance(x,A) \doteq Is-a(x,A)\\
&(InstA2)\ \exists x\ Instance(x,A)\ (every\ class\ A\ has\ some\ member\ x)\\
&(InstA3)\ \exists A\ Instance(x,A)\ (every\ individual\ is\ a\ member\ of\ some\ class)\\
&\\
&Instatiation\ relation\ as\ Subsumption\ relation\\
&\\
&Is-a(A,B) \doteq \forall x(Instance(x,A) \rightarrow (Instance(x,B))\\
&(every\ instance\ of\ A\ is\ also\ instance\ of\ B)
\end{aligned}
\tag{5.1}
$$

A spatial object is an object of spatial type in SOA1. SOA11 and SOA12 state that *AnatomicalEntity* and *MicroscopicEntity* are spatial objects.
Similarly, a spatial relation is a relation whose type is spatial in SRA1. The definitions SRA11 to SRA14 state that *SpatialRelation* subsumes all *MereotopologicalRelation*, *MetricRelation*, *GeometricRelation* and *DimensionRelation*. *SorroundnessRelation* is then defined as a *MereotopologicalRelatio*n (SRA111), *Distance* as *MetricRelation* (SRA121) and *Size* as *GeometricRelation* (SRA131).

$$
\begin{aligned}
&(SOA1)\ SpatialObject \sqsubseteq Object\ \Pi\ \exists type.Spatial\\
&(SOA11)\ AnatomicalEntity \sqsubseteq SpatialObject\\
&(SOA12)\ MicroscopicEntity \sqsubseteq SpatialObject\\
&\\
&(SRA1)\ SpatialRelation \sqsubseteq Relation\ \Pi\ \exists type.\ Spatial\\
&(SRA11)\ MereotopologicalRelation \sqsubseteq SpatialRelation\\
&(SRA12)\ MetricRelation \sqsubseteq SpatialRelation\\
&(SRA13)\ GeometricRelation \sqsubseteq SpatialRelation\\
&(SRA14)\ DimensionRelation \sqsubseteq SpatialRelation\\
&\\
&(SRA111)\ SuroundnessRelation \sqsubseteq MereotopologicalRelation\\
&(SRA121)\ DistanceRelation \sqsubseteq MetricRelation\\
&(SRA131)\ Size \sqsubseteq GeometricRelation
\end{aligned}
\tag{5.2}
$$

One consideration needs to be done with respect to two of these relations. The metric relations are special relations in the sense that they need to be defined in correspondence to a reference system, similarly as with spatial objects. Similarly to [Hudelot et al., 2008], we add two additional definitions for the spatial object and for the spatial relation.

$$
\begin{aligned}
&(SOA2)\ SpatialObject \sqsubseteq Object\ \Pi\ \exists type.Spatial\\
&\Pi\ \exists hasObject.ReferenceObject\\
&(SRA2)\ SpatialRelation \sqsubseteq Relation\ \Pi\ \exists type.\ Metric\\
&\Pi\ \exists hasSystem.ReferenceSystem
\end{aligned}
\tag{5.3}
$$

The other special category is represented by the geometric relations. Although geometric relations have been discussed in the literature from the qualitative perspective [Cohn and Renz, 2008], in our case, due to the nature of our ontological representation and its purpose in connecting the semantic level with the image analysis level, we do not provide any further definition in the formal theory of the relationships that are part of this category. This fact has to deal with the image analysis algorithms which provide efficient support for handling geometric relations such as size, shape or intensity, etc. We detail this aspect in chapter 8, in which we build our cognitive microscope framework.

5.1.2. Mereo-topological Relations

To better understand how exactly we apply this theory to breast cancer grading, it is necessary to present some of the definitions, axioms and theorems from [Donnelli et al., 2005]. Apart from the fact that mereology does not require numerical quantities, it is also to be noted that it does not require mathematical abstractions either (e.g. points, lines). It is on this underlying qualitative nature that all relations are build. A mereo-topology is generally an extension of mereology that includes also topological relations, such as connection, interior, and boundary, by abstracting size, shape or distance.

Different approaches on mereo-topology have been proposed. One worth mentioning is the region connection calculus theory also cited by [Donnelli et al.,

2005]. However, since our purpose is to extend Donnelli et al's work, we apply the region connection calculus to the metric relation in the next section.

One example relevant to us is the located in relation, based on *Loc-In* relation between individuals and on R_1 (A, B), R_2 (A, B), R_{12} (A, B) definitions of relations among classes.
The symbols used in definitions and formulas can be found in Table 2.7.

Definition 5.2. ***Located in*** (***Loc-In***) [Donnelli et al., 2005] is the mereological relation between two individuals x and y that holds whenever x's spatial region is located in y's spatial region, equivalent to say that x's location is included in y's location.

$$Loc-In(x,y) \doteq \mathrm{Pr}(x)r(y) \tag{5.4}$$

where Pr(x)r(y) indicates a parthood relation between x's and y's regions and not on x and y. An object x may be located in y but not part of y, although parthood relation drives the location relation. If x is part of y (Pxy), then x's region is part of y's region (Pr(x)r(y)) and is noted as:

$$Pxy \rightarrow \mathrm{Pr}(x)r(y) \tag{5.5}$$

Similarly, if two individuals x and y overlap (where O represents the overlap relation), it can be said that they partially coincide (Pcoin(x,y)).

$$(PCoinA1)\ PCoin(x,y) \doteq O\mathrm{r}(x)r(y) \tag{5.6}$$

The location relation can be also used to help describing some situation in which, x and y partially coincide, but do not overlap. The location axioms are noted as:

$$\begin{aligned} &(Loc-In_1)\ \ Loc-In(x,x)\ \ \textit{reflexive} \\ &(Loc-In_2)\ Loc\text{ - }In(x,y)\ \wedge\ Loc\text{ - }In(y,z) \rightarrow Loc\text{ - }In(x,z)\ \textit{transitive} \\ &(Loc\text{ - }In_3)\ \ \ Loc\text{ - }In(x,y) = Loc\text{ - }In(y,x)\ \rightarrow x = y\ \textit{antisymmetric} \end{aligned} \tag{5.7}$$

The first axiom says that object x is located in itself; (Loc-In_2) tells that if x is located in y and y is located in z, then x is located in z. The last axiom, (Loc-In_3), tells us that if x is located in y and y is located in x, then x and y are identical.

To apply *Loc-In* from individuals to classes, the use of the existential (some), universal quantifiers (every) and instantiation relation, are necessary to get the following theorems [Donnelli et al., 2005]:

$$
\begin{aligned}
&(LT1)\ Loc\text{ - }In_1(A,B) \doteq \forall x(Inst(x,A) \rightarrow \exists y(Inst(y,B) \wedge Loc\text{ - }In(x,y)) \\
&(LT2)\ Loc\text{ - }In_2(A,B) \doteq \forall y(Inst(y,B) \rightarrow \exists x(Inst(x,A) \wedge Loc\text{ - }In(x,y)) \\
&(LT3)\ Loc\text{ - }In_{12}(A,B) \doteq Loc\text{ - }In_1(A,B) \wedge Loc\text{ - }In_2(A,B)
\end{aligned}
\tag{5.8}
$$

Loc-In$_1$ (A, B) holds if and only if *every* instance of A is located in *some* instance of B.
Loc-In$_2$ (A, B) holds if and only if *every* instance of B *has some instance* of A located in it and
Loc-In$_{12}$ (A, B) holds when both are verified.

Note that this primary *Loc-In* used alone describes a generic location. There are some more specific relations (or location with additional conditions) that need to be defined. Hence, based on *Loc-In* definition from [Donnelli et al., 2005], we propose our definition for surroundness relation, which is relevant in the context of breast cancer grading spatial relations between spatial objects.

Definition 5.3. Surroundness (SurrBy) [Tutac et al., 2010a-b] is the mereological relation that holds between two individuals x and y whenever x's spatial region is located in y's spatial region but x and y do not overlap.
Based on the Loc-In(x, y) and ~Oxy definitions (localization and overlapping relations) from [Donnelli et al., 2005], we can derive a *Surroundness* mereological relation as follows [Tutac et al., 2009b], [Tutac et al., 2009c]:

$$
\begin{aligned}
&(SA1) SurrBy(x,y) \doteq \Pr(x)r(y) \wedge \sim PCoin(x,y)) \\
&or \\
&(SA2)\ SurrBy(x,y) \doteq Loc\text{ - }In(x,y) \wedge \sim O(x,y)
\end{aligned}
\tag{5.9}
$$

Concerning the logical properties of set operations for *SurrBy* relation, one could state the following axioms, related to the individuals.

$$
\begin{aligned}
&(SA3)\ \ SurrBy(x,x)\ \ (every\ individual\ is\ surrounded\ by\ itself)\ reflexive \\
&(SA4)\ SurrBy(x,y)\ \wedge\ SurrBy(y,z) \rightarrow SurrBy(x,z)\ transitive \\
&(SA5)\ \ SurrBy(x,y) = \ SurrBy(y,x)\ \rightarrow x = y\ antisymmetric
\end{aligned}
\tag{5.10}
$$

Similarly with Loc-In theorems, SurrBy$_1$ (A, B) holds if and only if *every instance* of A is surrounded by *some instance* of B. SurrBy$_2$ (A, B) holds if and only if *every instance* of B *has some instance* of A surrounded by it, and SurrBy$_{12}$ (A, B) holds when both.

$$
\begin{aligned}
&(ST1)\ SurrBy_1(A,B) \doteq \forall x(Inst(x,A) \rightarrow \exists y(Inst(y,B) \wedge SurrBy(x,y)) \\
&(ST2)\ SurrBy_2(A,B) \doteq \forall y(Inst(y,B) \rightarrow \exists x(Inst(x,A) \wedge SurrBy(x,y)) \\
&(ST3)\ SurrBy_{12}(A,B) \doteq Loc\text{ - }In_1(A,B) \wedge Loc\text{ - }In_2(A,B)
\end{aligned}
\tag{5.11}
$$

These defined relations help in disambiguating the representation and provide a reliable reasoning result. One illustration of disambiguation through elimination of redundant relation is by showing that an *Included-In* relation is not necessary since there is a *Located-In* and *SurrBy* relations already defined.
If we want to formulate the *Inclusion*, let use the Loc-In(x, y) in combination with PCoin [Donnelli et al., 2005] as follows:

$$(IA1)\ Included-In(x,y) \doteq \exists z Loc-In(x,z) \wedge Loc-In(y,z) \wedge \sim PCoin(x,y)) \quad (5.12)$$

An inclusion relation is therefore a relation between two individuals x and y if x's spatial region is located in y's spatial region and they do not partially coincide. Unless there is another relation in which overlapping does not mean partially coincidence, an Included-In relation is unnecessary.

5.1.3. Metric Relations

As specified in the generic definitions, in our approach the metric relations deal with the direction and distance information. In other representations they are viewed as part of topology. It is to be mentioned that they are dependent on each other, regardless of the distinction made on various approaches. However, it is difficult to define these relations without having a numerical support; they need to be defined in relation to a reference object or a reference system.
There are various ways to define the direction or distance relation, for instance the cardinal representation, or the trapezoidal rule [Hudelot et al., 2008]. However, based on the fact that [Stocker and Sirin, 2009] proposed a spatial qualitative approach for Pellet reasoner combining RCC-8 and OWL reasoning, we rely on RCC-8 composition table to define the *CloseTo* relation.
The 8 region relations are represented as the set {DC, EC, PO, TPP, TPPI, NTPP, NTPPI, EQ} between two regions x and y, where the regions contain objects or individuals. Figure 5.2 gives a visual representation of them:

- disconnected (DC)
- externally connected (EC)
- partially overlapping (PO)
- tangential proper part (TPP)
- tangential proper part inverse (TPPi); in the figure representing TPP, if y is instead of x, the relation becomes TPPi
- non-tangential proper part (NTTP)
- non-tangential proper part inverse (NTPPi) – y instead of x
- equal (EQ)

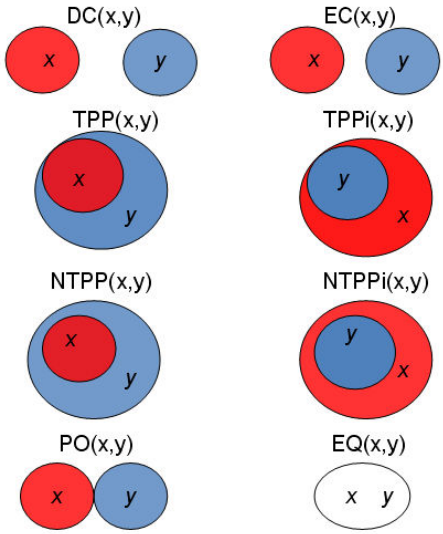


Figure 5.2. RCC-8 region calculus

Based on these 8 basic relations, a composition table can be constructed. The composition table shows the relation between region x and z, where R (x, y) and S (y, z), R and S being any of the 8 relations.

o	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ
DC	*	DC,EC,PO,TPP,NTPP	DC,EC,PO,TPP,NTPP	DC,EC,PO,TPP,NTPP	DC,EC,PO,TPP,NTPP	DC	DC	DC
EC	DC,EC,PO,TPPi,NTPPi	DC,EC,PO,TPP,TPPi,EQ	DC,EC,PO,TPP,NTPP	EC,PO,TPP,NTPP	PO,TPP,NTPP	DC,EC	DC	EC
PO	DC,EC,PO,TPPi,NTPPi	DC,EC,PO,TPPi,NTPPi	*	PO,TPP,NTPP	PO,TPP,NTPP	DC,EC,PO,TPPi,NTPPi	DC,EC,PO,TPPi,NTPPi	PO
TPP	DC	DC,EC	DC,EC,PO,TPP,NTPP	TPP,NTPP	NTPP	DC,EC,PO,TPP,TPPi,EQ	DC,EC,PO,TPPi,NTPPi	TPP
NTPP	DC	DC	DC,EC,PO,TPP,NTPP	NTPP	NTPP	DC,EC,PO,TPP,NTPP	*	NTPP
TPPi	DC,EC,PO,TPPi,NTPPi	EC,PO,TPPi,NTPPi	PO,TPPi,NTPPi	PO,TPP,TPPi,EQ	PO,TPP,NTPP	TPPi,NTPPi	NTPPi	TPPi
NTPPi	DC,EC,PO,TPPi,NTPPi	PO,TPPi,NTPPi	PO,TPPi,NTPPi	PO,TPPi,NTPPi	PO,TPP,NTPP,TPPi,NTPPi,EQ	NTPPi	NTPPi	NTPPi
EQ	DC	EC	PO	TPP	NTPP	TPPi	NTPPi	EQ

Table 5.1 The RCC-8 composition table [w3reg]

To understand how the composition table functions, let us consider the relations DC(x,y) and EC(y,z) and see what relation hold between x and z. Figure 5.3 shows that the following relations hold : DC, EC, TPP, NTPP and PO.

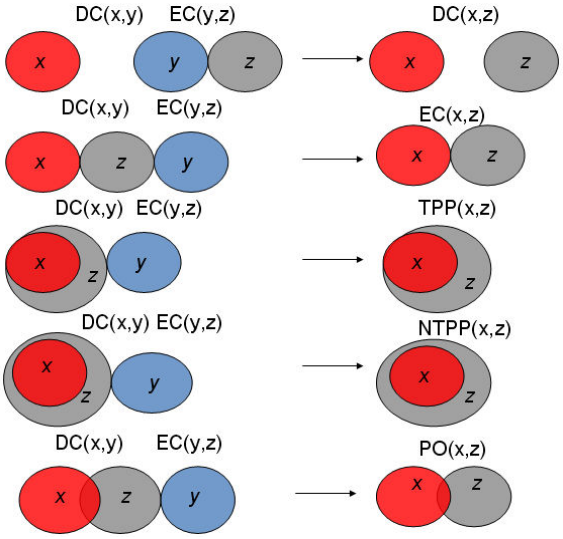


Figure 5.3. Composition between DC and EC relations

A comprehensive introduction to region connection calculus models and composition table is given by [Li and Ying, 2003]. All relations from the composition table can be similarly represented. As we stated for the surroundness relation, the relation between two individuals holds in fact between their spatial regions. In this light, we define the generic *CloseTo* relation as:

Definition 5.4. Closeness (CloseTo) is the metric relation that holds between two individuals x and y whenever x's spatial region (which is identical to x in this case) is disconnected to y's spatial region (which is identical to y in this case), or their spatial regions are externally connected, and there is another region z to which x's spatial relation is tangential proper-part and y's spatial region is externally connected to z's spatial region. The z region is viewed as a reference object, x and y are close to each other with reference to z [Tutac et al., 2010b].

$$(CA1)CloseTo(x,y) \doteq DC(x,y) \vee EC(x,y) \wedge TPP(x,z) \wedge EC(y,z), \quad (5.13)$$
where z<*threshold*

The condition set on region z, refers to the dimension of the region. For reasons of simplicity, we view the region as a circle with radius r. In this case, the radius r needs to follow the condition of being less than the threshold (r< threshold). The threshold depends on the exploration scale and values can be assigned to it in the SWRL module of our application ontology. We illustrate the closeness relation in Figure 5.4.

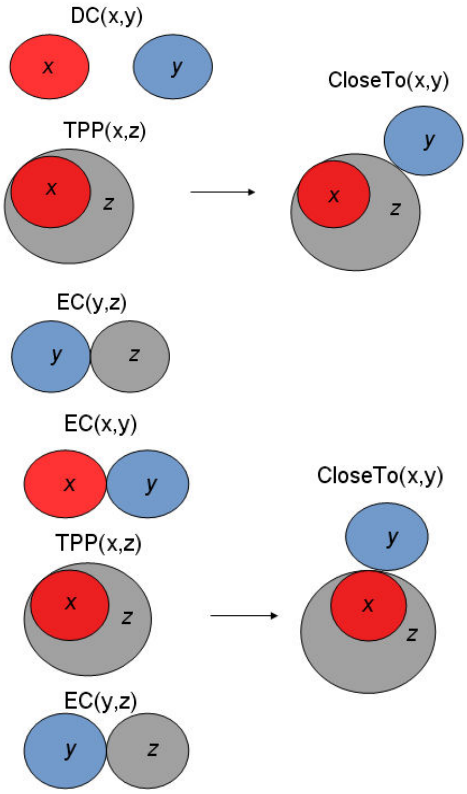


Figure 5.4. *CloseTo* relation

In a metric space, the following axioms need to be satisfied:

$$\begin{aligned}
&(CA2)\ CloseTo(x,y) \geq 0\ non-negativity \\
&(CA3)\ CloseTo(x,y) = CloseTo(y,x)\ symmetry \\
&(CA4)\ CloseTo(x,z) \leq CloseTo(x,y) + CloseTo(y,z)\ triangle\ inequality \\
&(CA5)\ CloseTo(x,y) = 0\ iff\ x = y\ identity
\end{aligned} \tag{5.14}$$

In the same manner, based on the RCC-8 relations and with the parameter specification from the medical experts, a *FarFrom* relation or a combination of distance relations (e.g. x is *CloseTo* y but *FarFrom* w) can be defined as shown in the approach of [Brageul and Guesgen, 2007].
Other distance relation such as left to, or orientation information such as anterior, posterior are not further detailed here, since a histopathologic assessment does not carry much relevant information related to these concepts, as a radiological exam does.

5.1.4. Dimension Relations

Scale dimension requires more complicated formulae, but we can simply illustrate the formal concept by extending the definition of a spatial region of an object x of [Grenon and Smith, 2004], which means that spatial region of x is the space of x at all scales.

$$(ScA1)\ Spatial\ Region(x) \doteq Part-of(x, space, res) \tag{5.15}$$

The axioms and theorems for the spatial region can be defined in a similar manner to which we defined the axioms and theorems for the parthood relation. We will illustrate this dimension relation in the following section.

5.2. Spatial Reasoning

The spatial reasoning can be performed in two ways: manual reasoning and automated reasoning. **Manual reasoning** represents the human inference process. In terms of the BCGO building methodology steps, the manual reasoning corresponds to the pathologist's feedback of the knowledge refining [Tutac et al., 2010a]. **Automated reasoning** corresponds to the Pellet reasoner which deals with computable logic consequences of the representation based on a DL reasoning algorithm. It is important to note that the DL reasoning handles all representations not only spatial representations. This section addresses the manual reasoning, while the next discusses the DL reasoning.

Since the ontology contains classes (*TB*ox) and individuals (*AB*ox), it is necessary to discuss the application of theorems to classes and/or individuals. It is also noted that both manual reasoning and automated reasoning apply not only to spatial concepts but to all formal representation from our model.

Let us consider the following example from BCG in which R1 (A, B) and R2 (A, B) both hold for mereo-topological ***surroundness*** relation.

$SurrBy_1$ (Lumina, StringNuclei)
Every Lumina is surrounded by *some StringNuclei* (holds true in conformity with the BCG domain knowledge) and

$SurrBy_2$ (Lumina, StringNuclei)
Every StringNuclei has *some Lumina* surrounded by it (holds true)

Therefore, $SurrBy_{12}$ (*Lumin*a, StringNucle*i*) holds.
(Every *Lumina* is located in *some StringNuclei* and *every StringNuclei* has *some Lumina* surrounding it and they do not overlap). Figure 5.5 shows these relations.

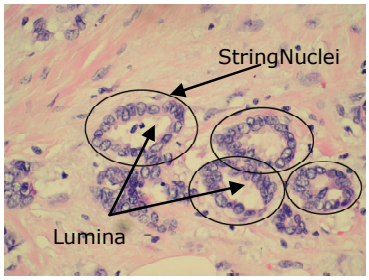


Figure 5.5. $SurrBy_{12}$ (Lumina, StringNuclei) on a microscopic image

Another illustration in which R_1 (A, B) holds but R_2 (A, B) does not hold in BCG, is for a ***partial-coincidence*** relation between *Nucleus* and *Mitosis* classes.

$PCoin_1$ (Mitosis*, Nucleus*)

Every Mitosis partially coincides with *some Nucleus* - holds

$PCoin_2$ (Mitosis, Nucleus)
Every Nucleus has *some Mitosis* partially coinciding with it – false, since not all nuclei are dividing and forming mitotic figures.

$PCoin_{12}$ (Mitosis, Tubule) is false since $PCoin_2$ is not holding true.
A similar situation is encountered for ***proper-parthood*** relation between *Cell* and *Nucleus* (only PP_1 (*Nucleus*, *Cell*) holds).

On the classes, all relations $SurrBy_1$, $SurrBy_2$ and $SurrBy_{12}$ are reflexive. Every Lumina is surrounded by some Lumina and every Lumina has some Lumina surrounded by it. In other words, every Lumina is surrounded by itself.
Symmetry instead can not be proven since it does not hold true for all situations. For instance, $SurrBy_1$ (Lumina, StringNuclei) states that every lumina is surrounded by some StringNuclei. This is not equal to say $SurrBy_1$ (StringNuclei, Lumina)- every StringNuclei is surrounding some Lumina (string nuclei are not located in lumina).
In terms of transitivity, the BCG domain knowledge does not offer us much information for the surroundness relation as it can only be proven for same type of relation – all $SurrBy_1$ or all $SurrBy_2$.

Considering the ***Loc-In*** relation, we could claim $Loc\text{-}In_1$ (InvasiveFrame, ROI) which holds since every invasive frame is located in some region of interest (ROI). The region of interest is detected according to an automated classification technique for the image processing. Our scope is to illustrate the reasoning with the spatial concept on the semantic level; therefore we will not delve any further into the image processing techniques involved. These techniques are described in detail in [Dalle et al., 2008], [Dalle et al., 2009], [Veillard et al., 2010], [Huang et al., 2010].
We can also infer that $Loc\text{-}In_2$ (InvasiveFrame, ROI) as every ROI has some invasive frames located in it. Hence $Loc\text{-}In_{12}$ (InvasiveFrame, ROI) holds true.

However, Loc-In_1 (ROI, Slide) is true, but we cannot apply the Loc-In_2 (ROI, Slide) as not every slide has some ROI located in it; there could be slides with no invasive area labelled as ROI.
To verify the transitivity property, we construct a Loc-In relation for classes, as in the following example.

Loc-In_1 (LargeNucleus, InvasiveFrame)
(Every large nucleus is located in some invasive frame) and

Loc-In_1 (InvasiveFrame, ROI)
(Every invasive frame is located in some ROI)

It follows logically that:

Loc-In_1 (LargeNucleus, ROI)
(Every large nucleus is located in some ROI).

The Loc-In relation between Slide, InvasiveFrame, ROI and LargeNucleus could be illustrated as in Figure 5.6:

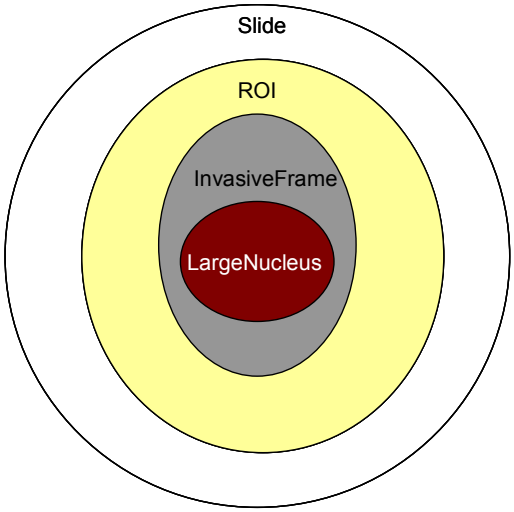


Figure 5.6. Loc-In relation for Slide (WSI), ROI, InvasiveFrame and LargeNucleus

The Whole Slide Imaging (WSI) represents the digitized slide which is formed of frames. The frames are also noted as High Power Fields (HPF) due to the fact that they are acquired at high magnification (40X). We will discuss more about WSI in chapter 8.

The following example shows the Loc-In relation on a microscopic image (see Figure 5.7)

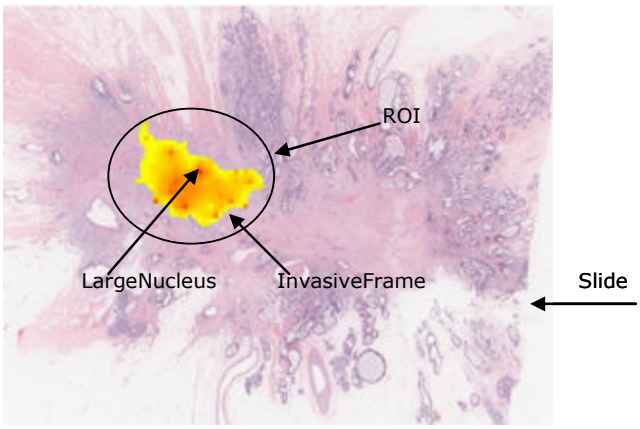


Figure 5.7. Loc-In relations on a microscopic slide (WSI)

In Figure 5.7, the following Loc-In relations are holding:
Loc-In_1 (LargeNucleus, InvasiveFrame),
Loc-In_1 (InvasiveFrame, ROI),
Loc-In_1 (ROI, Slide),
Loc-In_1 (LargeNucleus, ROI),
Loc-In_2 (LargeNucleus, InvasiveFrame),
Loc-In_2 (InvasiveFrame, ROI),
Loc-In_2 (LargeNucleus, ROI),
Loc-In_{12} (InvasiveFrame, ROI)
Loc-In_{12} (LargeNucleus, InvasiveFrame)
Loc-In_{12} (LargeNucleus, ROI)

Reflexivity and antisymmetry could be verified for classes for Loc-In_1 and Loc-In_2, Loc-In_{12} implying. Every LargeNucleus, InvasiveFrame, ROI and Slide are located in themselves (e.g. LargeNucleus_1, Slide_1). Therefore reflexivity axiom holds for location relation.
Similar to the surroundness relation, antisymmetry property of location relation cannot be proven. For instance Loc-In_2 (ROI, Slide) is false, as Loc-In_2 (Slide, ROI) ≠ Loc-In_2 (ROI, Slide) – it is true that every ROI has some ROI located in it but is not true that every ROI has some Slide located in it. In point of fact, no ROI has some Slide located in it.

More complex relations to reason with, could be a combination of defined relations applied to classes, as shown further:

$SurrBy_1$ (Lumina, StringNuclei)
(Every *Lumina* is surrounded by some *StringNuclei*)

P_2 (StringNuclei, Tubule)
(Every *Tubule* has some *StringNuclei* as a part)

We can infer:

SurrBy$_1$ (Lumina, Tubule)
(Every *Lumina* is surrounded by some *Tubule*)

We can also infer:
SurrBy$_2$ (Lumina, Tubule)
(Every *Tubule* has some *Lumina* surrounded by it)

If we consider the same R relations among classes for CloseTo concept, as shown, we can apply the following for the CloseTo (Mitosis, NeoplasmPeriphery).

$$\begin{aligned}&(CT1)\ CloseTo_1(A,B) \doteq \forall x(Inst(x,A) \rightarrow \exists y(Inst(y,B) \wedge CloseTo(x,y))\\ &(CT2)\ CloseTo_2(A,B) \doteq \forall y(Inst(y,B) \rightarrow \exists x(Inst(x,A) \wedge CloseTo(x,y))\\ &(CT3)\ CloseTo_{12}(A,B) \doteq CloseTo_1(A,B) \wedge CloseTo_2(A,B)\end{aligned} \tag{5.16}$$

CloseTo$_1$ (Mitosis, NeoplasmPeriphery) holds.
(Every instance of *Mitosis* is close to *NeoplasmPeriphery)*

CloseTo$_2$ (Mitosis, NeoplasmPeriphery) does not hold.
(Every instance of NeoplasmPeriphery has some instance of *Mitosis* close to it)
This statement is not true, since not every instance of neoplasm periphery presents mitosis close to it and since not all frames have mitosis. Hence CloseTo$_{12}$ (Mitosis, NeoplasmPeriphery) does not hold either. In Figure 5.8, CloseTo$_1$ (Mitosis, NeoplasmPeriphery) is illustrated on a microscopic image.

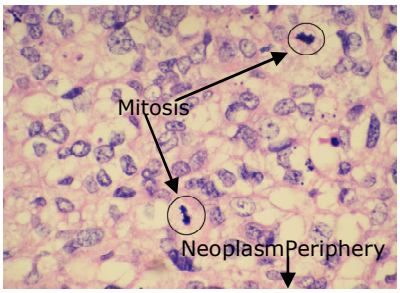


Figure 5.8. CloseTo$_1$ (Mitosis, NeoplasmPeriphery) on microscopic image

Let us verify the axioms of CloseTo relations for this particular case from breast cancer medical knowledge. Every *Mitosis* is close to NeoplasmPeriphery and the distance between Mitosis and NeoplasmPeriphery is greater than 0, therefore we can affirm that CloseTo$_1$ is non-negative. Mitosis is not NeoplasmPeriphery and vice-versa therefore the distance between them is different then 0 and identity axiom is holding too. In terms of symmetry, Close$_1$ (Mitosis, NeoplasmPeriphery) is not symmetric with Close$_1$ (NeoplasmPeriphery, Mitosis), since not every NeoplasmPeriphery is close to some Mitosis.

Similarly, Close$_2$ (Mitosis, NeoplasmPeriphery) ≠ Close$_2$ (NeoplasmPeriphery, Mitosis), every NeoplasmPeriphery has some instance of mitosis close to it is not symmetric with every mitosis has some instance of NeoplasmPeriphery close to it. In conclusion, we cannot say Close relations are symmetric. The triangle inequality

is hard to verify since breast cancer grading domain knowledge does not provide any meaningful information in this direction.

In the case of Mitosis, $SpatialRegion_{12}$ (Mitosis) = Part-of (Mitosis, neoplasm, 40X) does not hold, since it states that every instance of mitosis is part of some neoplasm at 40X resolution and every instance of neoplasm has some mitosis as part of it, at 40X resolution. The $SpatialRegion_2$ does not hold again.
Instead, $SpatialRegion_{12}$ (Nuclei) = Part-of(Nuclei, neoplasm, 10X) is true, since every instance of nuclei is part of some neoplasm at 10X resolution and every instance of neoplasm has some nuclei as part of it, at 10X resolution.
A similar reasoning could be applied to relations among all classes for the spatial concept introduced.

5.3. DL Reasoning

5.3.1. Tableau-based Algorithm

For the last generations, several algorithms have been developed to perform DL reasoning [Baader et al, 2003], [Baader et al, 2007].
A reasoning algorithm that can be used for consistency checking is the *structural subsumption algorithm* which compares syntactic structures of concept descriptions. However, this algorithm is useful only when little expressivity is needed (simple languages). In this case, the completeness of the reasoning is achieved.
The standard reasoning algorithm used currently in DL is the *tableau subsumption algorithm*, the 4^{th} generation of the DL reasoning method, allowing full negation and disjunction [Horrocks, 2000], [Baader and Sattler, 2000]. In this case, concept subsumption implies concept satisfiability. There are however many issues with regard to tableau-based algorithm or DL in general, such as the time complexity of the algorithm, which is a topic in itself and it is out of the scope of this current work. It is hence mentioned as future work in chapter 9 and discussions on it are presented in [Baader et al., 2007] and [Baader and Sattler, 2000].
This algorithm is implemented in the Pellet reasoner, the tool we use for performing the reasoning in BCGO.
The tableau-based calculus [Baader et al., 2007], [Baader and Sattler, 2000] takes the assumptions (concept subsumption, concept equivalence, etc) and makes a goal from deciding if their inverses are satisfiable. For instance:

$$C \subseteq D \text{ iff } C \sqcap \neg D = \varnothing \tag{5.17}$$

Concept C is subsumed by D iff the intersection of C with the complement of D is empty.
Before starting the tableau algorithm, several preliminary steps are applied to the expressions targeted to be satisfiable proved.

1. ***Unfolding or TBox elimination*** which implies a recursive expansion of all concepts to their definitions from the knowledge-base until only primitive concepts are in the formula. Instead of proving all $K \sqcap C \sqcap \neg D \models \varnothing$ we replace C

and D with regard to K, we prove $U_1 \Pi \neg U_2 \models \varnothing$, where U_1 is the unfolded C and U_2 is the unfolded D.

2. ***Normalization*** means application of Negation Normal Form of De Morgan to the formula, where the negation only applies to concept names and not to compound terms.

$$\begin{aligned} \neg \exists R.C &\equiv \forall R.\neg C \\ \neg \forall R.C &\equiv \exists R.\neg C \\ \neg \leq nR.C &\equiv \geq (n+1)R.C \\ \neg \geq (n+1)R.C &\equiv \leq n.R.C \\ \neg \geq OR.C &\equiv \mathrm{C} \prod \neg \mathrm{C} \end{aligned} \tag{5.18}$$

Given an unfolded and normalized formula, the tableau algorithm starts building a tree-like model of an input concept. Hence the goal of the algorithm is to try to construct a model of the formula. The concept is syntactically decomposed by *applying tableau expansion rules* and inferring constraints on the elements of the models. When there are no more rules to apply or whenever clash (an obvious contradiction) occurs, the algorithm stops. A concept is satisfiable if rules can be applied such that a fully expanded clash-free tree is constructed. More specifically:

Nodes represent individuals. Edges represent properties of individuals.
Each node *x* is labeled with a set of concept expressions and it must satisfy $\alpha(x) = \{C_1, \ldots, C_n\}$
Each edge <*x, y*> satisfies a role and is labeled with its name $\alpha(< x, y >) = R$
Given an expression *E*, a tree *T* is initialized to contain a single node x_0, with $\alpha(x_0) = \{E\}$
T is expanded by repeatedly applying tableau rules (shown in Figure 5.9 and Figure 5.10).
A branch is closed when for a node *x* and some concept *C*, either $\{C, \neg C\} \in \alpha(x)$ or $- \in \alpha(x)$

The ***tableau expansion rules*** handle the elimination of the four types of operators: the conjunction elimination, the union elimination, the existential and the universal elimination [Koubarakis, 2010].

$$
\begin{array}{lll}
\sqcap-rule & if\ 1. & (C_1 \sqcap C_2) \in \mathcal{L}(x) \\
 & 2. & \{C_1, C_2\} \not\subseteq \mathcal{L}(x) \\
 & then & \mathcal{L}(x) \longrightarrow \mathcal{L}(x) \cup \{C_1, C_2\} \\
\sqcup-rule & if\ 1. & (C_1 \sqcup C_2) \in \mathcal{L}(x) \\
 & 2. & \{C_1, C_2\} \cap \mathcal{L}(x) = \emptyset \\
 & then & \text{a. save } \mathbf{T} \\
 & & \text{b. try} \mathcal{L}(x) \longrightarrow \mathcal{L}(x) \cup \{C_1\} \\
 & & \quad \text{If that leads to a clash then restore } \mathbf{T} \text{ and} \\
 & & \text{c. try} \mathcal{L}(x) \longrightarrow \mathcal{L}(x) \cup \{C_2\} \\
\exists-rule & if\ 1. & \exists R.C \in \mathcal{L}(x) \\
 & 2. & \text{there is no } y \text{ s.t. } \mathcal{L}(\langle x, y \rangle) = R \text{ and } C \in \mathcal{L}(y) \\
 & then & \text{create a new node } y \text{ and edge } \langle x, y \rangle \\
 & & \text{with } \mathcal{L}(y) = \{C\} \text{ and } \mathcal{L}(\langle x, y \rangle) = R \\
\forall-rule & if\ 1. & \forall R.C \in \mathcal{L}(x) \\
 & 2. & \text{there is some } y \text{ s.t. } \mathcal{L}(\langle x, y \rangle) = R \text{ and } C \notin \mathcal{L}(y) \\
 & then & \mathcal{L}(y) \longrightarrow \mathcal{L}(y) \cup \{C\}
\end{array}
$$

Figure 5.9. The tableau expansion rule [Herchenröder, 2006]

A concept is satisfiable if a possible *AB*ox is found applying the tableau-based algorithm. If no *AB*ox is found under all the search paths, then the concept is not satisfiable.
The tableau algorithm has to meet three requirements [Baader et al., 2007], [Koubarakis, 2010]:

- Soundness – if a complete and clash-free *AB*ox is found by the algorithm, the *AB*ox satisfies the initial concept (an *AB*ox is complete if none of the expansion rules apply to it);
- Completeness – if the initial concept is satisfiable, the algorithm can always find a complete and consistent *AB*ox (an *AB*ox is consistent if no logic clash is found);
- Termination– the algorithm can terminate in finite steps with specific results.

Any instantiation must comply with the constraints of the ontology. Given a statement from the ontology, the role of semantics is to logically decide by means of

interpretation what the *models* of the statement are. That is, to find all possible instantiations of the domain, compatible with the specified statement, hence to prove the statement is true. Several models are built by the ontology in the process of inference. The intersection of all models of each statement from the ontology represents the set of models which are carried out when the reasoning is complete.
For example, the ontology states that *Nucleus* is a subclass of *MicroscopicEntity* (in any possible situation, each nucleus is also a microscopic entity) and if it is known that *Nucleus_1* is a *Nucleus* (*Nucleus_1* is an instance of *Nucleus* class) then, in any possible situation it is necessarily true that *Nucleus_1* is a *MicroscopicEntity*, since the situation in which it would not be a *MicroscopicEntity* is not compatible with the constraints.

Another example viewed from the steps of the tableau algorithm perspective, is that of Mitosis being subsumed by MicroscopicEntity [Tutac et al., 2010a]. Giving the following axioms and assertions from the knowledge-base:

$$Nucleus \sqsubseteq MicroscopicEntity$$
$$NucleusDivision \sqsubseteq Nucleus$$
$$Mitosis \sqsubseteq NucleusDivision \sqcap \neg\forall isLocatedIn.Tubule$$
$$Mitosis\_1 \sqsubseteq Mitosis$$

The query goes as following: Is *Mitosis_1 subsumed by MicroscopicEntity?* In DL terminology, that is to prove unsatisfiability of:

$$Mitosis\_1 \sqsubseteq \neg MicroscopicEntity$$

1. ***Unfolding*** step firstly relies on the subsumption of *Mitosis_1* by *Mitosis*, and therefore on replacing the concept name *Mitosis* with its definition expansion. The *Mitosis* is further replaced by *NucleusDivision*, which at its turn is replaced by *Nucleus* and finally it reaches the top axiom, *MicroscopicEntity*.

$$Mitosis\_1 \cup Mitosis \equiv NucleusDivision \sqcap \neg\forall isLocatedIn.Tubule$$
$$Mitosis\_1_1 \sqcap NucleusDivision \sqcap \neg\forall isLocatedIn.Tubule$$
$$Mitosis\_1_1 \sqcap NucleusDivision_2 \sqcap Nucleus \sqcap \neg\forall isLocatedIn.Tubule$$
$$Mitosis\_1_1 \sqcap NucleusDivision_2 \sqcap Nucleus_3 \sqcap MicroscopicEntity \sqcap \neg\forall isLocatedIn.Tubule$$
$$Mitosis\_1_1 \sqcap MicroscopicEntity \sqcap \neg\forall isLocatedIn.Tubule$$

2. ***NNF*** applies to the last transformed formula which now becomes :

$$Mitosis\_1_1 \sqcap MicroscopicEntity \sqcap \exists isLocatedIn.\neg Tubule$$

3. ***Proving with tableau-based algorithm***

$$\mathcal{L}(x) = \{MicroscopicEntity \sqcap \exists isLocatedIn.\neg Tubule\}$$

Our label node $\mathcal{L}(x)$, where the individual x is *Mitosis_1* contains the formula given by the NNF transformation. As we have conjunction expression in the statement we apply the conjunction rule elimination first, followed by the existential elimination (Figure 5.10).

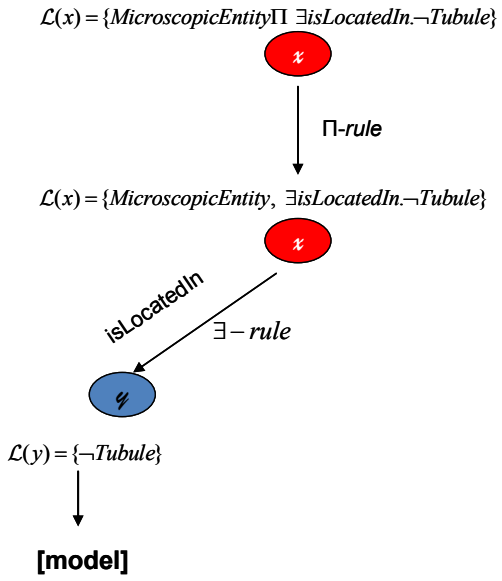


Figure 5.10. Subsuming *Mitosis_1* by *MicroscopicEntity*

The algorithm finished in a finite amount of time, and the concept was found satisfiable, therefore we can say that the *Mitosis_1* is subsumed by *MicroscopicEntity*. This also means that we found an interpretation $\mathcal{I}$ which satisfies the assertion from the $\mathcal{L}(x)$. Notice that the spatial relation *isLocatedIn* is also part of the description of the *Mitosis_1*; hence the reasoning verified the spatial assertion as well.

5.4. Conclusions

In this chapter we presented our solution to build a consistent ontology for breast cancer grading domain in order to bridge the semantic gap.
We introduced a formal theory for spatial representation of breast cancer grading spatial mereo- topological, metric and dimension relations (e.g. *SurrBy*, *Loc-In*, *CloseTo*) extending the mereological theory of [Donnelli et al., 2005] or using RCC-8 and composition table in line with the integration of spatial aspect to the DL reasoner recently proposed by [Stocker and Sirin, 2009]. With a formal theory support we strengthen the modeling of domain knowledge and we help in alleviating

ambiguities in the image interpretation as we showed for *Loc-In* and *Inclusion* relations.
The consistency of ontology has been verified by two means: manual reasoning and DL reasoning.
In the manual reasoning mode, we showed how to apply the spatial axioms and theorems in breast cancer grading domain as the medical reasoning naturally works. Not all axioms of set relations could be proven for the statements of BCG. Unlike reflexivity, in most of the cases, antisymmetry property did not hold. Transitivity was best illustrated with *Loc-In* relation.
For some metric relationships and for dimension relation, we verified the specific distance relation axioms, such as symmetry and non-negativity. The triangle inequality could not be verified due to lack of medical information support in this direction.
In the DL reasoning, we explained the steps taken by the DL reasoner in order to build a model of the input assertions and thus to verify the satisfiability of them. It is important to mention that the DL reasoner takes all assumptions, including the spatial assumptions and verifies them. In other words, the spatial reasoning is part of the DL reasoning.
In closing this chapter, a summary of the contributions of the thesis at this point is given:

- Formal spatial theory for the representation of breast cancer grading spatial concepts and relations as a formal framework of combined mereo-topological, geometrical and metrical spatial relations. This theoretical framework helps in reducing the ambiguities and inconsistencies in representation.
- the spatial representation follows the formulation of qualitative desideratum
- two approaches for reasoning with spatial relations in BCG : manual by relying on the resemblance with the medical reasoning and automated based on DL tableau algorithm

6. Model Implementation. Breast Cancer Grading Ontology

This chapter is dedicated to the BCGO implementation. We follow the knowledge translation and knowledge refining steps from the BCGO model design. The BCGO is implemented using Protégé[4] framework, an open source ontology editor and knowledge acquisition system developed in Java by Stanford Centre for Biomedical Informatics Research and we work with the Pellet[5] tool [Sirin et al., 2005] integrated into Protégé for the reasoning process.

According to the DL formalism, the knowledge-base is composed of *TB*ox and *AB*ox. We populate these boxes, showing the correspondence between them and OWL classes, properties and instances, in section 6.1. We discuss the characteristics of OWL in implementation and various modalities of formal representations. The next section contains the modeling of the SWRL rule in various situations. We construct the SWRL module based on our thesis regarding qualitative versus quantitative representations for semantic concepts, spatial knowledge included as well. Computational aspects are treated from the DL safety standpoint. The BCGO stands as the single ontology for breast cancer grading aiming at integration into higher reference ontologies

6.1. DL *TB*ox. OWL Classes and Constraints

The *TB*ox contains the definitions of concepts and the statement of constraints, defined generically as axioms, whilst the *AB*ox contains the assertions: the concept and role/properties assertions. The *TB*ox corresponds to the OWL classes and properties and the *AB*ox contains the instances of classes and properties.
To synthesize and highlight the main differences in feeding the knowledge-base $\mathcal{K}$, a parallel between *TB*ox and *AB*ox is given in Figure 6.1 [Roux et al., 2009a]. In this example, the nucleus is a microscopic entity which has some size *Size*. *Nucleus_1* is an instance of *Nucleus* whose size is *SmallSize*.

$\mathcal{TB}$ox	$\mathcal{AB}$ox
definitions of concepts $Nucleus = MicroscopicEntity \sqcap ..$	concept assertions Nucleus (Nucleus_1)
statement of constraints $\exists hasSize.Size \subseteq Size$	roles assertions hasSize(Nucleus_1, SmallSize)

Figure 6.1. *TB*ox & *AB*ox

[4] http://protege.stanford.edu/, last accessed July 2010
[5] http://clarkparsia.com/pellet, last accessed July 2010

We start building our ontology by first defining the medical core concepts – viewed and translated as classes in the *Owl:Thing* hierarchy. We classify them into four main categories: *AnatomicalEntity*, *ConceptualEntity*, *MicroscopicEntity* and *SpatialEntity,* respectively.

The first category defines anatomical concepts such as *Breast, Tissue,* in order to have a link with the anatomical description of breast from the NIH thesaurus. The key point is to be able to connect concepts from this category with concepts from our specific domain, for instance the Ducts where the *Disease* occurs (Ductal Carcinoma in Situ).
The purpose of the second category is to handle generic concepts from the world of grading, connecting processes such as *Assessment* with objects as *Specimen* and *Person* or *Disease.* In other words, *ConceptualEntity* class contains structured information with respect to persons (*Person*) that are patients (*Patient*) having a disease (*Disease*), which need to be assessed (*Assessment*), using specific criteria (*Criterion*) of NGS (*NottinghamGrading*) on a specimen (*Specimen*) analysis. Note that when we discuss about *Assessment,* even though it is viewed as a process in medical terms, the result is given at a specific moment in time and in our representation we capture only a static moment (according to our time-independent approach).
The third class contains the concepts related to microscopic objects from the histopathology images, objects characteristic to grading of breast carcinoma. Hence, *MicroscopicEntity* describes NGS relevant objects identified on a specimen analyzed under the microscope – cell, nucleus, mitosis, tubule, lumina, etc. Figure 6.2 illustrates the structure of this class.
In the same way, the fourth family of concepts relates to the spatial descriptors on an image, with the purpose of connecting the high level concepts with the low level concepts from the image presented in *MicroscopicEntity*. It unfolds into three types of descriptors: *Mereo-TopologicalDescriptor, MetricDescriptor* and *GeometricDescriptor* respectively.
In the first class and second class we have the parthood, location, distance and direction concepts as discussed in the spatial theory. The *GeometricDescriptor* structures features of objects mainly in terms of *Size*, *Shape* and *Intensity,* since these features are essential to the pathologist reasoning in performing the grading over frames or slide.
Note that we do not introduce any image processing detailed information with respect to size, shape or intensity in the ontology. We only structure these concepts in order to guide the image processing phase.
Based on the table from refining phase we tackle the most important characteristics of OWL-DL language in ontology implementation such that errors are avoided.

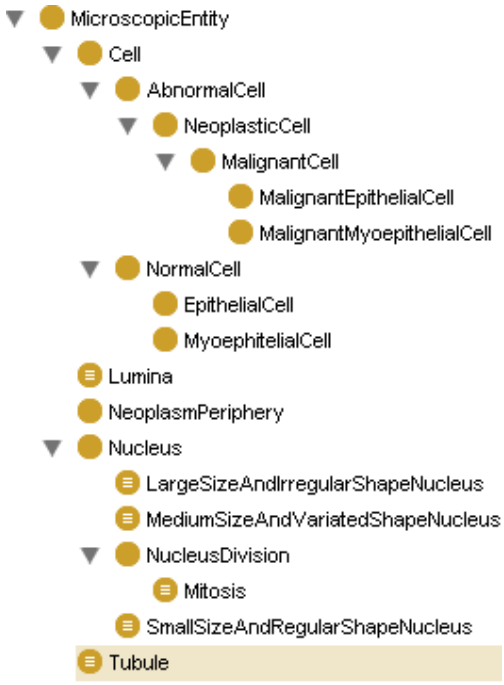


Figure 6.2. MicroscopicEntity fragment

6.1.1. Defined Classes and Primitive Classes

One of the novel characteristics of OWL classes stands in the difference between defined classes and primitive classes and coming with that, the possibility to convert from one to another.
In OWL abstract syntax, the defined classes are called complete classes and primitive classes are called partial classes. A primitive class presents a concept by a set of necessary conditions. Classes for which there is at least one set of necessary and sufficient conditions fall into the defined classes' category. An alternative to name defined classes is equivalent classes, using the equivalence symbol. This is highly important due to how the mechanism of reasoning handles with them, for the classifier generally infers nothing to be subsumed under a primitive class except of domain or range constraints. It is therefore of great significance to make a definition complete or necessary and sufficient, rather than partial or just necessary.
According to [Rector et al., 2004], there are three basic principles on which one should decide which class to make defined or let primitive: the pragmatic reason, the philosophical and the timing heuristics.

Let consider the following example taken from the Concept class, for the *CriteriaScoring*. The *NuclearPleomorphismScoringOne* class is defined as in Figure 6.3:

OWL:

Class (NuclearPleomorphismScoreOne complete NuclearPleomorphismScoring
restriction (hasNucleus someValuesFrom SmallSizeAndRegularShapeNucleus)
restriction (hasNucleus allValuesFrom SmallSizeAndRegularShapeNucleus))

OWL-DL:

$NuclearPleomorphismScoreOne \equiv NuclearPleomorphismScoring \sqcap$
$\exists hasNucleus.SmallSizeAndRegularShapeNucleus \sqcap$
$\forall hasNucleus.SmallSizeAndRegularShapeNucleus$

Figure 6.3. *NuclearPleomorphismScoreOne* class in OWL and OWL-DL

To paraphrase, a nuclear pleomorphism score one is *any* nuclear pleomorphism scoring that *amongst other things*, has *only* small size and regular shape nucleus. A formulation like *amongst other things* allows a *NuclearPleomorphismScoreOne* to have other restrictions than *hasNucleus, but* these two restrictions are necessary and sufficient to define any nuclear pleomorphism score one. We want things to be classified under this class automatically; hence it fulfills the pragmatic principle. From the philosophical point of view, the nuclear pleomorphism score one can be defined completely, even though the pathologists sometimes look for other details within the tissue. But if the nuclei are small in size and their shape is rather regular, it is sufficient for them to score it as one. Lastly, there is no solid justification to let this definition of the concept nuclear pleomorphism score one, primitive at this moment in time and change it later to complete.
The use of both quantifiers - someValuesFrom and allValuesFrom, will be further explained under another characteristic of OWL, namely the Open World Assumption (OWA). However, it is important to mention here also, that in OWL language, allValuesFrom does not imply someValuesFrom.

An example of a primitive class is the one of *Nucleus (*Figure 6.4*)*:

OWL:
Class (Nucleus partial MicroscopicEntity
restriction (isLocatedIn someValuesFrom Cell)
restriction (hasSize someValuesFrom Size)
restriction (hasSize allValuesFrom Size)
restriction (hasShape someValuesFrom Shape)
restriction (hasShape allValuesFrom Shape))

OWL-DL:
$Nucleus \sqsubseteq MicroscopicEntity \prod$
$\exists isLocatedIn.Cell \prod$
$\exists hasSize.Size \prod \forall hasSize.Size \prod$
$\exists hasShape.Shape \prod \forall hasShape.Shape$

Figure 6.4. *Nucleus* primitive class in OWL and OWL-DL

What we say in fact is that nucleus is a microscopic entity that is *amongst other things located in some cell, also has size only Size and also has shape only Shape*. However if some individual satisfies these conditions we can not say that it is a member of class *Nucleus*, even though if an individual is a member of this class then it must satisfy these conditions.
We let the nucleus concept primitive for the moment, with the possibility to make it a defined class, as some other restrictions may be added in the future, if for instance, some other characteristic that the pathologist will look for will be essential in the definition.

Another facet of complete classes versus partial classes is when some restrictions do not transform into necessary and sufficient conditions but remain as necessary implications. These necessary implications require a conversion to subclass axioms which in its turn implies syntactic changes.
For instance, a restriction which states that a Frame which has *Disease DuctalCarcinomaInSitu* or *LobularCarcinomaInSitu* only *NuclearPleomorphismScoring* is taken into consideration, is not a necessary and sufficient condition, but rather is further needed to be logically inferred whenever something is found to be a *Frame (*Figure 6.5*)*.

OWL:
Class (Frame partial VirtualSpecimen
restriction (hasDisease someValuesFrom Disease)
restriction (hasDisease allValuesFrom Disease)
restriction (hasNottinghamScoring someValuesFrom NottinghamScoring)
restriction (hasNottinghamScoring allValuesFrom NottinghamScoring))

would be syntactically changed into:

OWL:
Class (Frame complete VirtualSpecimen
restriction (hasDisease someValuesFrom Disease)
restriction (hasDisease allValuesFrom Disease)
restriction (hasNottinghamScoring someValuesFrom NottinghamScoring)
restriction (hasNottinghamScoring allValuesFrom NottinghamScoring))
SubclassOf (Frame
restriction (hasDisease only DuctalCarcinomaInSitu or LobularCarcinomaInSitu)
restriction (hasNottinghamScoring only NuclearPleomorphismScoring)))

Figure 6.5. *Frame* primitive versus defined class

6.1.2. Disjoint Classes and Subsumption Hierarchy

Another important implementation principle regarding classes in OWL is the disjointness concept. Classes are assumed to overlap if disjoint claims are not explicitly made, and an individual can be an instance of more than one class in the same time. Essentially, all siblings of a class need to be disjoint from one another, in order to obtain a consistent classification and representation of concepts.
To illustrate the principle of disjointness, let us analyze the Assessment class which has 2 subclasses: *HistopathologicalGrading* and *HistopathologicalScoring*.
These concepts are different in the medical terminology as well; the score obtained on each different criterion is used to give the final grading. Similarly, in our ontology, we cannot assume that an instance of *HistopathologicalScoring* is not a member of *HistopathologicalGrading* just because it has not been asserted to be an instance of this class. We need to make this statement explicit. We need to make sure that none of the instances of *HistopathologicalScoring* are instances of *HistopathologicalGrading*. Thus, the histopathological grading and histopathological scoring are explicitly made disjoint from one another.
In point of fact, a related common misunderstanding of OWL formalism mechanism is the failure to make all information explicit.

Given the following definition from Figure 6.6:

OWL:
Class (DuctalCarcinomaInSitu complete DuctalBreastCarcinoma
restriction (hasCell someValuesFrom MalignantCell)
restriction (hasCell allValuesFrom MalignantCell))

OWL-DL:
$DuctalCarcinomaInSitu \equiv DuctalBreastCarcinoma \prod$
$\exists hasCell.MalignantCell \prod$
$\forall hasCell.MalignantCell$

Figure 6.6. *DuctalCarcinomaInSitu* defined class in OWL

We want to say that the malignant cell can be either epithelial and/or myoepithelial cell, where the epithelial cells, according to the breast anatomy, are the inner layer of cell ducts and the myoepithelial cells are the outer layer, opposed to the basement membrane (the basal lumina). Hence, we capture this detail of information by adding the restriction of values from *MalignantCell* and only from *MalignantCell*.
Furthermore, we naturally assume that the ductal carcinoma in situ occurs in the ducts (it is localized in the ducts) according to the medical description, but the definition does not say it. Thus, in order for the reasoner to work with this information, an explicit description of it is required. The definition becomes as in Figure 6.7:

OWL:
Class (DuctalCarcinomaInSitu complete DuctalBreastCarcinoma
restriction (hasCell someValuesFrom MalignantCell)
restriction (hasCell allValuesFrom MalignantCell)
restriction (isLocatedIn someValuesFrom Ducts)
restriction (isLocatedIn allValuesFrom Ducts))

OWL-DL:
$DuctalCarcinomaInSitu \equiv DuctalBreastCarcinoma \prod$
$\exists hasCell.MalignantCell \prod$
$\forall hasCell.MalignantCell \prod$
$\exists isLocatedIn.Ducts \prod$
$\forall isLocatedIn.Ducts$

Figure 6.7. *DuctalCarcinomaInSitu* defined class with existential restrictions

To this end, one can see that this corresponds to the refining step from the BCGO modeling process and this aspect can be observed in other representations that will follow.

This principle of disjointness applies to every classification we explicitly define in our ontology (e.g. the MicroscopicEntity class contains *Nucleus, Tubule, Lumina, Cell and NeoplasmPeriphery* subclasses which are all disjoint from one another).
One could note that all siblings of a particular class are to be made disjoint. The classes *Tubule*, *Cell* and *Lumina* are disjoint to the selected class *Nucleus*. Also note that 'inherited disjointedness' is applying, such that although *EpithelialCell* or *MalignantCell* are not explicitly stated to be disjoint to *Nucleus*, they are in fact disjoint because their super classes are disjoint to *Nucleus.*

Another highly related issue to the OWL disjointness concept is the subtle issue of *is-a* relationship when developing a class hierarchy. Containment and subclass relation do not have to be confused. For instance, *Nucleus* is a component of a *Cell*, but is *not* a *Cell*. This comes to stress the need of a formal theory support for relations such as parthood and proper-parthood, (theory which we presented in chapter 5).

6.1.3. Open World Assumption

In the semantic web technologies, Open World Assumption (OWA) is an important feature yet difficult to understand when it comes to negation restriction [Rector et al., 2004], [Drummond and Shearer, 2006]. In OWL, negation is seen as unsatisfiability unlike in constraint languages, logic programming, etc, which follow a close world assumption approach where this is considered as failure.
Essentially, the close world assumption implies that everything we do not know (that is not represented in the description) is false, whilst the open world assumption approach implies that everything we don't know is undefined or not computable and is found to be false only if there is a proof that it contradicts some representation from the ontology.

Let us consider the following fragment of the tubule concept definition (Figure 6.8):

OWL:
Class (Tubule complete MicroscopicEntity
restriction (hasCell allValuesFrom (NormalCell or MalignantCell)))

Figure 6.8. *Tubule* class in OWL

The definition says that a tubule is any microscopic entity that has only *NormalCell* or *MalignantCell*. We consider the *Tubule* class is complete; these conditions are necessary and sufficient for any microscopic entity to be a tubule.

Given now a *CarcinomaTubule* definition (Figure 6.9):

OWL:
Class (CarcinomaTubule complete MicroscopicEntity
restriction (hasCell someValuesFrom (EpithelialCell or
MalignantEpithelialCell or MalignantMyoepithelialCell)))

Figure 6.9. *CarcinomaTubule* class in OWL- OWA issue

This states that a carcinoma tubule has amongst other things epithelial cells or malignant epithelial cells or malignant myoepithelial cells, as some cell may still be normal while others are malignant.
We would assume that a carcinoma tubule is a tubule based on the subsumption relation between *Tubule* and *CarcinomaTubule*, due to the fact the cells of a carcinoma tubule, defined by set of necessary and sufficient conditions, are subclasses of *Normal* and/or *MalignantCell*.
However when we run the classifier, we notice that the *CarcinomaTubule* is not classified under *Tubule*.
This means there is something missing in the definition. In point of fact, the existential quantifier *someValuesFrom* used for *hasCell* property for the representation of *CarcinomaTubul*e, gives a possibility for the reasoner to assume that a carcinoma tubule can have some other kinds of cells, other than *EpithelialCell* or *MalignantEpithelialCell* or *MalignantMyoepithelialCell*.
We solve this issue by adding what is called a **closure axiom** on the hasCell property. A closure axiom consists of a universal restriction that given to the property states that it can only be filled by the specified fillers. A carcinoma tubule can have *EpithelialCell* or *MalignantEpithelialCell* or MalignantMyoepithelialCell and *only* these fillers.

The correct definition of CarcinomaTubule is illustrated by Figure 6.10:

OWL:
Class (CarcinomaTubule complete MicroscopicEntity
restriction (hasCell someValuesFrom (EpithelialCell or
MalignantEpithelialCell or MalignantMyoepithelialCell))
restriction (hasCell allValuesFrom (EpithelialCell or
MalignantEpithelialCell or MalignantMyoepithelialCell)))

Figure 6.10. *CarcinomaTubule* class correct definition with closure axiom

The mitosis definition provides another example of closure axioms requirement. In OWL-DL, the mitosis definition is formalized in the following way (Figure 6.11):

OWL:
Class (Mitosis complete NucleusDivision
restriction (hasIntensity someValuesFrom VeryLow)
restriction (isCloseTo someValuesFrom NeoplasmPeriphery)
complementOf (restriction (isLocatedIn someValuesFrom Tubule)))

Figure 6.11. *Mitosis* defined class with OWA

According to this definition mitosis is any microscopic entity that has amongst other things some very low intensity and also is close to some neoplasm periphery and also is not located in some tubule. However when creating an instance of this class, the restriction on has intensity property says that there is at least one relationship has intensity with *VeryLow,* but a mitosis can have other intensities with has intensity property as well. In other words, there is nothing in the definition that limits the intensity only to very low, as the medical terminology would require in order to identifying mitosis. Therefore, we must add a closure axiom on the hasIntensity property.
A mitosis can have intensity *VeryLow* and only *VeryLow*. Similarly, a closure axioms applies to *isCloseTo* and *isLocatedIn* property, because a mitosis is close to *NeoplasmPeriphery* and only to *NeoplasmPeriphery* (it can not be found else where in the tissue) and is not located only in tubule. Unless these restrictions are explicitly specified, mitosis can have very high intensity as well, or could be/not be located in other place than the tubule.
The definition now is transformed as it is depicted by Figure 6.12.

A remark is to be mentioned here. ***The OWA does not refer to the complementOf***- the negation restriction, although OWA may accompany a negation restriction as well, as we previously shown in the mitosis definition (the complementOf restriction was use to capture the information that the mitosis is not located in the tubule, according to the NGS). The negation is unsatisfiability means that we cannot assume that something does not exist until is explicitly said that it does not exist.

Apart from the closure axiom that restricts the existential statement with a universal statement, it is worth mentioning the opposite action.
In the tubule definition, we wanted to say that a tubule is any microscopic entity that has only *NormalCell* or *MalignantCell*. However, setting only the universal restriction without a corresponding existential restriction along the hasCell property merely signifies that the Tubule individuals only participate in hasCell relationship with members from *NormalCell* or *MalignantCell*, and also those individuals that do not participate in any hasCell relationships. This leads to an error, thus the existential restriction should be added in order to have a correct representation and a complete class; these both conditions then are necessary and sufficient for any microscopic entity to be a tubule.

OWL:

Class (Mitosis complete NuclearDivision
restriction (hasIntensity someValuesFrom VeryLow)
restriction (hasIntensity allValuesFrom VeryLow)
restriction (isCloseTo someValuesFrom NeoplasmPeriphery)
restriction (isCloseTo allValuesFrom NeoplasmPeriphery))
complementOf (restriction (isLocatedIn someValuesFrom Tubule))
complementOf (restriction (isLocatedIn allValuesFrom Tubule)))

OWL-DL:

$Mitosis \equiv NuclearDivision \prod$
$\exists hasIntensity.VeryLow \prod \forall hasIntensity.VeryLow \prod$
$\exists isCloseTo.NeoplasmPeriphery \prod$
$\forall isCloseTo.NeoplasmPeriphery \prod$
$\neg\exists isLocatedIn.Tubule \prod \neg\forall isLocatedIn.Tubule$

Figure 6.12. *Mitosis* defined class with universal restriction - closure axiom

To conclude this section, we point out in synthesis some of the reasons of OWA's applicability problems:
A reason what the Open World Assumption is difficult to understand and apply consists of ***the trivial satisfiability of only (allValuesFrom) restrictions***. In OWL language, *allValuesFrom* does not imply *someValuesFrom*. To identify the superficial satisfiable restrictions early, an existential restriction (*someValuesFrom*) is added to correspond to every universal restriction (*allValuesFrom*) either in the class or in one of its subclasses.

Additionally, a ***misunderstanding or an incorrect usage of universal and existential quantifiers*** on properties leads also to not be able to apply the OWA and therefore to obtain different classification than a human reasoning would expect and produce.
The existential restriction describes at least one relationship that holds between a given property and an individual that is a member of a specific class (called the filler).
For instance, $\exists hasIntensity\ VeryLow$ or (*hasIntensity some VeryLow*) restricts individuals from *Mitosis* to have at least one relationship to an individual from *VeryLow* class. It does not imply that all the relationships *hasIntensity* must be of the class VeryLow. In order to restrict the instances of Mitosis class to have *hasIntensity* relationship with only V*eryLow* members of the *VeryLow* class, the universal restriction must then be used.
A universal restriction constrains the relationship to individuals that are member of a specific class. Saying $\forall hasIntensity\ VeryLow$ or (*hasIntensity all VeryLow*) connects the member of *Mitosis* via hasIntensity property to only member of *VeryLow* class (all of whose hasIntensity relations are to *VeryLow* class instances).

Formulated in another way, the individuals of *Mitosis* are not allowed to have other *hasIntensity* relationship other than with *VeryLow* members.
In line with that, another subtle problem concerns ***the difference between logical and linguistic use of unionOf (or) and intersectionOf (and).*** Let us analyze one of the previous definitions (Figure 6.13):

OWL:
Class (CarcinomaTubule complete MicroscopicEntity
restriction (hasCell someValuesFrom (EpithelialCell or
MalignantEpithelialCell or MalignantMyoepithelialCell))
restriction (hasCell allValuesFrom (EpithelialCell or
MalignantEpithelialCell or MalignantMyoepithelialCell)))

Figure 6.13. *CarcinomaTubule* defined class- unionOf and intersectionOf

The meaning of this definition is that for an individual to be a member of class *CarcinomaTubule* is necessary and sufficient to only have cell that are *EpithelialCell or MalignantEpithelialCell or MalignantMyoepithelialCell* (*either* individuals from *EpithelialCell or* from the other two classes).
Using an "and" operator instead of an "or", would turn in saying the hasCell property is hold to individuals that are simultaneously *EpithelialCell* and *MalignantEpithelialCell*, etc., which would be logically incorrect. Now, *EpithelialCel*l is disjoint to all mentioned classes. Hadn't been so, it would have sounded naturally logic and the reasoner would have not triggered any inconsistency.

The next situation presented in Figure 6.14 is more subtle:

OWL:
Class (LargeSizeAndIrregularShapeNucleus complete Nucleus
restriction (hasSize someValuesFrom LargeSize)
restriction (hasSize allValuesFrom LargeSize)
restriction (hasShape someValuesFrom IrregularShape)
restriction (hasShape allValuesFrom IrregularShape))

Figure 6.14. *LargeSizeAndIrregularShapeNucleus* – unionOf and intersectionOf

In agreement with the NGS description of a nucleus that has large size and irregular shape, we reflect in our formalized definition that a *LargeSizeAndIrregularShapeNucleus* is *any Nucleus* whose size is *LargeSize* and *only LargeSize* and whose shape is *IrregularShape* and only *IrregularShape*. When creating an individual from this class, the only allowed value for size is *LargeSize* and the only allowed value for *Shape* is *IrregularShape*.
Notice that there is no unionOf or intersectionOf explicitly written. However, when multiple restrictions are used, the total description takes the intersection of the individual restriction. And in this case there are no contradictions, since we want to

say that a *LargeSizeAndIrregularShapeNucleu*s individual is simultaneously having *LargeSize* and *IrregularShape.* Hence, the restrictions are simultaneously true, also because the restrictions are directed to two different aspects –the size and the shape, but they need to be fulfilled in the same time.

6.2. DL *AB*ox. OWL Properties and Instances

In the *AB*ox, individuals of classes are instantiated; for example *Slide* individuals or *Frame* individuals can be introduced and connected through the *hasAssessment* relationship with the *Patient* individuals.

One of the ways of representing individuals in a class definition is by using enumeration feature.
Let us consider the following example given in Figure 6.15:

OWL:
Class (Slide complete VirtualSpecimen
restriction (hasIdentifier someValuesFrom Identifier)
restriction (hasIdentifier allValuesFrom Identifier)
restriction (hasNottinghamGrading someValuesFrom NottinghamGrading)
restriction (hasNottinghamGrading allValuesFrom NottinghamGrading)
restriction (enumeration {Slide1 Slide2}))

Figure 6.15. Slide enumerated class (nominals)

The Slide class is called an enumerated class and an individual that is a member of Slide class must be one of the individuals specified in the brackets. *Slide1* individual *hasIdentifier* NB50752007 and *hasNottinghamGrading GradeTwo_1*.The list may be filled further while developing the ontology and creating instances of Slide based on the information provided by the medical experts.

The relationships in our ontology are defined as to formally represent the relations from the medical terminology. We need both types of properties from OWL: the object and the datatype properties. Object properties link one individual to another individual, while datatype properties link one individual to a data literal whose type may be string, integer, float, etc. There is another category of properties, the annotation properties namely, but they deal with some predefined properties which we do not explicitly manipulate such as protégé: allowed Parent, protégé: abstract, etc.
From the DL standpoint, the object and datatype properties may also be identified as role assertions and they thus fall into the *AB*ox representation. In OWL terms, there is no clear distinction between statement of constraints and role assertions, as the statement of constraints lead eventually to instances and the relationships

among them. For the sake of clarity, we present them connected with the definition of concepts.
The object properties as well as the datatype properties may form hierarchies of sub-properties, in line with the hierarchy of classes. It is common understanding that a datatype property cannot have an object property as sub-property and vice versa. However, it is not possible to mixture them, there is no option when creating them that allows the use to do that.

6.2.1. Object Properties

One simple example of an object property is *hasAssessment* or *hasDisease*.
A breast cancer patient must have an assessment performed in order to analyze the disease status. If he or she is a breast cancer patient also has a disease. For instance, *BreastCancerPatient_*1 *hasDisease DuctalCarcinomaInSitu_8* and *has Assessment NottinghamGrading* and this could be even more specifically the *NottinghamGrading GradeOne_1*. Both properties are specializations of the super-property *hasConceptualEntity*.

Object properties may have corresponding **inverse properties**. One such example that is very relevant to the breast cancer grading domain we model is the *SurroundedBy* object property which links individuals from *Lumina* class to individuals from *Tubule* class. Its inverse is *isSurrounding* - individuals from *Tubule* surrounding instances of *Lumina*.
Another example is for hasIdentifier and its inverse isIdentifierOf: *Patient_1 hasIdentifier Identifier_1*, which would imply that *Indetifier_1 isIdentifierOf Patient_1*.
The following example shows that not all relations that seem to have an inverse that we would think of actually have it.
The object property *isCloseTo* links *Mitosis* individuals to *NeoplasmPeriphery* individuals. If *isFarFrom* stands as the inverse of *isCloseTo*, it would imply that NeoplasmPeriphery individuals are far from *Mitosis*. This statement is not true, in light to what we previously mentioned and also according to the NGS restrictions. Hence, *isFarFrom* is not the inverse of *isCloseTo*.

Object properties may have different characteristics such as transitive, functional, reflexive, symmetric, inverse functional, etc. We will only illustrate some of them for the relationships from our ontology.

For instance is*LocatedIn* is a **transitive property,** as also shown in the formal spatial theory.
If this property is transitive and individual *Nucleus_1 isLocatedIn Cell_1* and *Cell_1 isLocatedIn Frame_1*, we can infer that *Nucleus_1 isLocatedIn Frame_1*. Furthermore, *Frame_1 hasNucleus MediumSizeandVariatedShapeNucleus_1,* to be more precise.

If a property has **functional characteristic**, it links one individual to at most one individual via that property. One such example from our ontology is the *hasIdentifier* property. A patient can only have a single identifier; similarly a frame or a slide can only have a single identifier.

If we say that Slide_1 *hasIdentifier* NB50752007 and we further state that Slide_1 *hasIdentifier* NB50422007, because *hasIdentifier* is a functional property, the reasoner will infer that NB50752007 and individual NB50422007 is the same individual. This is also due to the mechanism of Open World Reasoning. A close world assumption would trigger an error, there can only be one identifier.
However, if these individuals are explicitly stated to be different from each other, this representation would definitely lead to inconsistency.
In point of fact the *hasIdentifier* is an **inverse functional property**, and its inverse, *isIdentifierOf* is functional.
Based on what we presented on transitive and functional characteristics, we can imply logically that a transitive property cannot be functional. Also, if a transitive property has an inverse correspondent the latter must also be transitive.
When talking about the *isCloseTo* property we conclude that its inverse could not be *isFarFrom*. Yet, this property has an inverse, but its inverse is exactly itself. What we say is that *isCloseTo* is **symmetric**. Giving the example with the mitosis and the neoplasm periphery again, we state that *Mitosis_1 isCloseTo NeoplasmPeriphery_1* and because the property is symmetric we can deduce that NeoplasmPeriphery_1 *isCloseTo Mitosis_1*.

6.2.2. Datatype Properties

Datatype properties are mostly having data literals as integer or boolean. The boolean literal is needed for *hasMitosis* property for instance, when we state that a *Slide* has *Mitosis* or not. This kind of information comes in line with the reasoning of the medical experts, as they are interested in firstly deciding if they go for a higher grade in their assessment or not.
The property *hasNumberOfFrames* is a datatype property that links individuals of *TenHyperFields* class with integer data literals. A *TenHyperFields* indicate a number of frames to be analyzed when counting for mitosis and this number is 10 frames, according to the NGS. Hence, we set a cardinality restriction to *hasNumberOfFrames* to exactly 10.
We also benefit of the possibility to make datatype property especially to handle the values for Nottingham scoring. The property *hasScoreBetween3And5* has allowed values 3, 4, and 5 and is used in the description of *ScoreBetween3and5* Class. Similarly for *SizeInPercentage*, we add allowed values, because the medical terminology is based on it when assessing the *TubularFormation.* Alternatively, *hasScoreBetween3and5* could be defined using a SWRL rule with min and max values.
Even though size naturally goes with numerical values, the reasoning we did not categorized *hasSize* or *hasIntensity* as datatype properties is firstly because the medical reference terminology does not specify what a small size means, or what very high intensity means. The pathologists operate with this qualitative knowledge and our approach is also qualitative, therefore we follow the same way. Furthermore, the formal semantic description gives possibility to the image processing level to interconnect and to take this property and work on it within the code.
Yet, if highly needed, some values could be given in the SWRL modules, and in this case the hasSize and hasIntensity must be of type integer. That does not mean the

qualitative principle is no longer holding. In our ontology, the qualitative principle stands in all situations, in the light of the qualitative versus quantitative perspective. The next section discusses further on other arguments with regard to choosing objects properties or datatype properties.

6.2.3. Object Properties or Datatype Properties

There are some situations in which it becomes difficult to choose what kind of property would be for some of the relationships. Thus, we further discuss the possible scenarios that may be encountered.
One illustration is the *hasNottinghamScoring* property which we firstly say that is an object property which links an individual of Frame class to an individual of *NottinghamScoring* class. Note the closure axiom to make sure that the *Nottingham scoring* of any frame has only values of *NottinghamScoring*.

The description is as following (Figure 6.16):

OWL:
Class (Frame complete VirtualSpecimen
restriction (hasNottinghamScoring someValuesFrom NottinghamScoring)
restriction (hasNottinghamScoring allValuesFrom NottinghamScoring))

Figure 6.16. *Frame* class- *hasNottinghamScoring* as object property

One could argue that this particular object property could better be a datatype property. In this case, the definition would look as depicted by Figure 6.17:

OWL:
Class (Frame complete VirtualSpecimen
restriction (hasNottinghamScoring someValuesFrom int))

Figure 6.17. *Frame* class- *hasNottinghamScoring* as datatype property

While this is for the reason that in natural language this sounds like a numerical value, note that in formal language it is much difficult to represent it, especially that the scoring in Nottingham Grading takes some particular values and not the entire range of integer values. It follows that this setting requires a decision on ***qualitative representation with quantitative values***.

If we state that *Frame_20 hasNottinghamScoring* 20 this would not be identified as inconsistency by the reasoner, as there was no restriction to specify the highest score possible according to NGS.
Additionally, while ***OWL has cardinality restrictions which gives the possibility to say that a property has a certain number of values (.e.g. 9 values), a datatype property cannot be restricted further, due to the connection with***

XML schema specification requirements. Nevertheless, saying that *hasNottinghamScoring* may have 9 values (9 being the highest score), in this specific case does not help at all, since this is not what we want to represent and even so, the values for scoring are having gaps in between.
An alternative would be to allow a user-defined integer range interval from score 3 (min) to score 9 (max). A module called xsp.owl that handle this feature could be imported to the ontology. Yet, one of the drawbacks is that the DL reasoner is not able to work with it.
Another possibility is to turn to a SWRL rule in which swrl builtins could express the min and max values. We will discuss this solution at the SWRL rules section.

Having said all that, as we want to have ***high expressivity without loosing decidability power,*** the only reliable solution is to make the *hasNottinghamScoring* property an object property, following that the *NottinghamScoring* class is composed of three subclasses: *ScoreBetween3And5*, *ScoreBetween6And7* and *ScoreBetween8And9*, each of which are defined using properties as *hasScoreBetween3and5* some int, and similarly to the others.

6.2.4. Property Domain and Range

The purpose of domain and ranges is to link individuals from a domain to individual from a range when necessary. In OWL-DL reasoning, the domain and ranges are viewed as axioms and not as constraints to be checked. This fact is of high significance due the multitude of inconsistency errors these axioms generate. They may cause a class to be unsatisfiable or a classification of a class under a complete other class than expected.

For instance, the individuals from *Nucleus* class or *Tubule* class are linked to individuals from *Size* via the *hasSize* property. The domain for *hasSize* property could be either *MicroscopicEntity* or owl:Thing (the universe) and the range is *Size*. If the domain would be set as to *MicroscopicEntity* this will lead to inconsistency, hence we choose *owl:Thing* as the domain. Similarly axioms apply for *hasShape* and *hasIntensity* property.

Another illustrative example is for *hasNottinghamScoring* properties. According to our description, *NottinghamGrading* instances are linked to individuals from *NottinghamScoring* via *hasNottinghamScoring* property. The domain then for *hasNottinghamScoring* is *NottinghamGrading* and the range is *NottinghamScoring*. However these axioms generate many classes from the ontology to be unsatisfiable.
For properties that are depicting relationships among different classes from different categories is not wise to limit to a domain and a particular range. The *isLocatedIn* is such a representative. *Nucleus isLocatedIn* Cell, where *Nucleus* and *Cell* are subclasses of *MicroscopicEntity*. A *DuctalCarcinomaInSitu* disease *isLocatedIn Ducts,* where *DuctalCarcinomaInSitu* is subclass of *DuctalBreastCarcinoma* and of higher super-class *Disease,* while *Ducts* are explicitly classified as *MicroAnatomicalEntity*. It does not make sense to limit the links between individuals by domain and range set to one of these classes. It is possible yet to specify multiple classes as range for a given property. In such a case, the range view would be interpreted as a union of the classes mentioned.

In the end of this section, is it important to say that there are other various aspects in the implementation of our BCG ontology which we did not extend any further. We consider that what we presented is neccessary and sufficient to clearly understand the characterstics of the OWL-DL formalism.

6.3. SWRL Rules

In the previous section we showed how to implement classes, properties and instances, in the light of the characteristics and issues of OWL-DL formalism. We also mentioned that there are different ways of formally representing a concept definition and restrictions and we pinpointed several situations in which alternatives of a different representations using SWRL would be another solution. SWRL module is an exemplifying representation of the qualitative with quantitative values combination.
One of the first questions that arise is when exactly to apply a rule? Are there always alternatives of representation?
Firstly, the context will tell whether it is possible or not. In our case, the context is the breast cancer domain. Secondly, a representation must follow the restrictions of the formalisms used. Lastly, there are situations when multiple representations (using only OWL or OWL combined with SWRL) are both correct and consistent and it is a matter of preferences and also of time computation; and there are situations in which the OWL is limited and the SWRL comes to meet the need of a complete description.
In this section, we address in more detail the SWRL approach based on our ontology model and on these remarks.

6.3.1. SWRL Rules Alternative to OWL

We showed the syntax of SWRL in chapter 2. To make it clear to understand, Table 6.1 provides examples from BCGO.

Let us bring forth one of the previous examples, concerning the property hasScoreBetween3and5.
In SWRL terms, this property would be represented as:

$$\text{ScoreBetween3and5}(?x) \wedge \text{hasScoreBetween3and5}(?x, ?value) \wedge \text{swrlb: greaterThan}(?value, 3) \wedge \text{swrlb: lessThan}(?value, 5) \rightarrow \text{hasScoreBetween3and5}(?x, ?value)$$

or alternatively, creating a relationship between hasScore and hasScoreBetween3and5 properties (predicates):

$$\text{ScoreBetween3and5}(?x) \wedge \text{hasScore}(?x, ?value) \wedge \text{swrlb: greaterThan}(?value, 3) \wedge \text{swrlb: lessThan}(?value, 5) \rightarrow \text{hasScoreBetween3and5}(?x, ?value)$$

SWRL symbols	Syntax	Semantics	Description
$\wedge$		conjunction	
$\rightarrow$		implication	
?a		variable	
	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$	C- class name i.e. Slide(?x)
	$R(a,b)$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$	R- object property; a, b- object variable name/object individual name i.e. hasMitoticScoring (?x,?mitoticscore)
	$U(a,v)$	$(a^{\mathcal{I}}, v^{D}) \in U^{\mathcal{I}}$	U- data type property v- data type variable name/data type value name i.e.MitoticScoring (?npm, 3)
	$D(v)$	$v^{D} \in \Delta^{D}$	D- data type
	$builtIn(p, v_1 \ldots v_n)$	$(v_1{}^{D}, \ldots v_n{}^{D} \in p^{D})$	p- built-in names i.e. swrlb: greaterThanOrEqual (?score,2)
	$a = b \;/\; a \neq b$	$a^{\mathcal{I}} = b^{\mathcal{I}} \;/\; a^{\mathcal{I}} \neq b^{\mathcal{I}}$	

Table 6.1. SWRL syntax in BCGO

Similarly, like we mentioned in arguing about Size and Intensity descriptors, if a specific value for defining a small size for the nucleus is necessary (e.g for testing), a SWRL rule looks like:

SmallSizeAndRegularShapeNucleus(?x) Λ hasSize(?x, ?value)
Λ swrlb: greaterThan(?value, 50) Λ swrlb: lessThan(?value, 100) →
hasSmallSize(?x, ?value)

At this point, it is noticed that with this kind of representation we only partially captured the characteristics of a small sized and regular shaped nucleus. It is therefore necessary to define it in a different yet complete way.

Nucleus (?x) Λ hasSize(?SmallSize, ?value) Λ hasShape (RegularShape, ?value) →
SmallSizeAndRegularShapeNucleus(?x)

This alternative does not assign numerical values to small size and regular shape, but attaches small size and regular shape to variable ?value in order to define a nucleus that is SmallSizeAndRegularShapeNucleus.
The following SWRL rule is for the hasNottinghamScoring property to capture the various values it could take according to the NGS.

$$\text{NottinghamScoring}(?x) \wedge \text{hasScore}(?x, ?value) \wedge \text{swrlb: greaterThan}(?value, 3) \wedge \text{swrlb: lessThan}(?value, 9) \rightarrow \text{hasNottinghamScoring}(?x, ?value)$$

With respect to spatial relations, one illustrative example that we already presented in our formal spatial theory is for *CloseTo* description, where we made use of a system to be referred to for defining closeness information.

6.3.2. Syntactic Sugar Rules

In chapter 2, we introduced the concept of syntactic sugar rules, which basically means a SWRL rule that can be translated into a DL restriction without the help of SWRL constructors. The translation from SWRL to DL depends on the number of variables based on the shared variables between the consequent and the antecedent.
If one grasps the implication of this idea, one would understand that this is an answer to the question as to when apply a SWRL rule and to make it efficiently.

Assume that we want to express such a rule:

$$\text{Nucleus}(?x) \wedge \text{hasMediumSize}(?x,?y) \wedge \text{hasVariatedShape}(?x,?z) \rightarrow \text{hasNuclearPleomorphismScoreTwo}(?x,?y)$$

The execution of this rule concludes in binding the *hasNuclearPleomorphismScoreTwo* property to y, in the individual that satisfies the rule, named x, where x is connected by variables y and z to properties *hasMediumSize* and *hasVariatedShape*.
As we see, this rule cannot be translated from SWRL to DL, for it has two variables shared between the antecedent and consequent.

An example of a syntactic sugar rule:

$$\text{Patient}(?p) \wedge \text{hasDuctalCarcinomaInSitu}(?p,?dcis) \wedge \text{isLocatedIn}(?dcis,?ducts) \wedge \text{Ducts}(?ducts) \rightarrow \text{BreastCancerPatient}(?p)$$

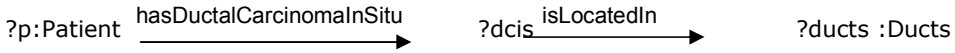
This rule can be expressed only using DL as following:

$$\mathit{BreastCancerPatient} \sqsubseteq \mathit{Patient} \sqcap \exists \mathit{hasDuctalCarcinomaInSitu}.\exists \mathit{isLocatedIn}.\mathit{Ducts}$$

A three step procedure is used in order to transform a SWRL rule into DL.
1. In the first step, **the consequent and the antecedent become two conjunctive queries**.

a. ?p: BreastCancerPatient

b. ?p: Patient(?p) Λ (?p,?dcis) : hasDuctalCarcinomaInSitu Λ (?dcis,?ducts) : isLocatedIn Λ ?ducts: Ducts
c. The cojunctive terms build a direct graph, in which each node is a variable of a named individual and each node is a relation. The query graph will then become:



2. The second step consists of **translating the resulted query into a class expression using a rolling-up technique**. Each edge is presented as a restriction; outgoing edges are transformed into an existential quantifier where expressions of the type: (?a, ?b):R become an expression of $\exists R.b$ where B is the class restriction on b.

$$\exists hasDuctalCarcinomaInSitu.\exists isLocatedIn.Ducts$$

The named class of the target variable ?p is given by Patient. The final result of the rolling up technique consist of intersection between this named class of the target variable and the rest of the expression.

$$Patient \sqcap \exists hasDuctalCarcinomaInSitu.\exists isLocatedIn.Ducts$$

3. Lastly, the antecedent becomes the subclass of the consequent

$$BreastCancerPatient \sqsubseteq Patient \sqcap \exists hasDuctalCarcinomaInSitu.\exists isLocatedIn.Ducts$$

6.3.3. SWRL Only

One particular kind of information from breast cancer grading domain cannot be expressed using OWL description: the sum between individual criteria scorings. The nuclear pleomorphism scoring is added to tubule formation scoring and added to mitosis count scoring to give the scoring.
In other words, that is precisely a situation in which we could only rely on SWRL, if we want to reflect this kind of information as well in our ontology.

NottinghamGrading (?x) Λ hasNottinghamScoring(?x, ?value)
Λ swrlb: sum (hasTubuleFormationScoring, hasMitosisCountScoring,
hasNuclearPleomorphismScoring, ?value) →
hasNottinghamScoring(?x, ?value)

6.3.4. Combining Ontology with Rules. SWRL DL Safe Rules

In the previous example, we showed that sometimes SWRL can express what OWL cannot.
Here, we point out a combination of OWL representation with SWRL representation to give a more complete "semantic picture" of what a concept from the BCG

knowledge mean. The purpose of combining ontologies with rules is to achieve higher expressivity.
One such illustrative example is the one of mitosis from Figure 6.18:

OWL:
Class (Mitosis complete NuclearDivision
restriction (hasIntensity someValuesFrom VeryLow)
restriction (hasIntensity allValuesFrom VeryLow)
restriction (isCloseTo someValuesFrom NeoplasmPperiphery)
restriction (isCloseTo allValuesFrom NeoplasmPeriphery))
complementOf (restriction (isLocatedIn someValuesFrom Tubule))
complementOf (restriction (isLocatedIn allValuesFrom Tubule))

OWL-DL:

$$Mitosis \equiv NuclearDivision \prod$$
$$\exists hasIntensity.VeryLow \prod \forall hasIntensity.VeryLow \prod$$
$$\exists isCloseTo.NeoplasmPeriphery \prod$$
$$\forall isCloseTo.NeoplasmPeriphery \prod$$
$$\neg\exists isLocatedIn.Tubule \prod \neg\forall isLocatedIn.Tubule$$

SWRL:

$$Nucleus(?x) \wedge hasEccentricity(?x,?value) \wedge$$
$$swrlb:lessThan(?value,1) \wedge swrlb:greaterThan(?value,0) \rightarrow$$
$$Mitosis(?x)$$

Figure 6.18. Mitosis defined class. OWL-DL and SWRL rules

In our BCGO model depicted by Figure 4.6, the SWRL rule for mitosis captures the hasEccentricity property, which in mathematical terms is a parameter for conic sections, the deviation from a circular shape. This interests us in capturing the characteristic of the changing in shape of the nucleus when diving (toward an ellipse). In this case, we create a restriction in OWL-DL on the hasEccentricity relation which connects to the SWRL module.

The example from above represents another alternative in which, the SWRL rule is separated from the OWL-DL module, but the reasoner is aware of it and infers according to this information as well. The rule now concerns not the dependence between two properties, but the definition of Mitosis in terms of eccentricity. Mitosis is any nucleus whose eccentricity value is between 0 and 1. The threshold can be set up higher as to catch only values between 0.5 and 1 if one considers that higher values than an irregular shape might give are needed.

As discussed in chapter 2 there are various aspects which concern the DL safety SWRL rules. Generally, to make a SWRL rule to be safe from the DL point of view,

thus meaning to offer decidability - which is the key desiderata- every variable from the consequent (the head) must also appear in the antecedent (the body). Note that

safety rule does not mean that every variable from the antecedent must appear in the consequent.
For instance, the rule for BreastCancerPatient is a safe-rule, the variable ?p from the head appears in the body of the rule.

$$\text{Patient}(?p) \wedge \text{hasDuctalCarcinomaInSitu}(?p,?dcis) \wedge \text{isLocatedIn}(?dcis,?ducts) \wedge \text{Ducts}(?ducts) \rightarrow \text{BreastCancerPatient}(?p)$$

When combining ontology with rules, it is possible that non-DL atoms or predicates are involed. A non-DL predicate is a Datalog atom with a predicate symbol that does not occur as a class or property in any OWL axiom. There are two diferent degrees of safety: strong and weak safety and a particular aspect that regards the role safety. As it is more expected to obtain decidability with strong-safety condition, rather than only with weak-safety (for the strong-safety implies the weak-safety), we show how strong-safety applies to our ontology. A strong-safety conditin means that variables from SWRL rules must be bounded only to explicitly defined individuals from the ontology.
For instance, a similar example with the one from above:

$$\text{hasDisease}(?p, ?dcis) \wedge \text{hasAssessment}(?p, ?dcis) \rightarrow \text{BreastCancerPatient}(?p)$$

Notice there is a difference between this statement and the previous definition which says that a breast cancer patient (?p) is a patient (?p) who has DuctalCarcinoma In Situ (?dcis) which DuctalCarcinomaInSitu is located in Ducts (?ducts).
What we say now is that hasDisease and has Assessment are bound to the same variable ?p and the assessment of the disease is bound to the same variable ?dcis. However, in this case, this rule is not a safe DL rule because hasDisease and hasAssessment both occur in OWL axioms.

Hence the solution is to enforce the DL-safety by restricting rules to named individuals.
The DL-safe rule is transformed to:

$$\text{hasDisease}(?x, ?value) \wedge \text{hasAssessment}(?x, ?value) \wedge \text{Patient}(?x) \wedge \text{NotthinghamGrading}(?value) \rightarrow \text{BreastCancerPatient}(?x)$$

6.4. Protégé framework and Pellet reasoner

As mentioned in the beginning of the chapter, we used Protégé framework to develop our BCGO ontology. Protégé is a free open source Java-based ontology editor and knowledge management environment. The main advantages of this tool are architecture flexibility and extensibility for rapid prototyping and application

development. It provides two ways of modeling ontologies: Protégé-Frames and Protégé-OWL. We modeled the BCGO in Protégé-OWL due to its characteristics of assisting in editing ontologies in OWL, of accessing DL reasoners and of querying instances for semantic indexing. In addition, Protégé-OWL is used by a large community of both academic and developers in a variety of domains (although it has been historically proposed for biomedical applications) and it is envisioned to become a standard infrastructure for building ontology-based semantic web applications [Knublauch et al., 2004]. A comprehensive presentation of the design, meta-model and features of Protégé-OWL as an extension of the Protégé core system is also given in [Knublauch et al., 2004]. We only illustrate the meta-model workflow in Figure 6.19 and we note that this meta-model is applicable to the 3.x series BCGO is developed in Protégé-OWL 3.4.4 build 579, OWL 1.0 and we plan to further migrate to Protégé 4.0 which uses OWL 2.0.

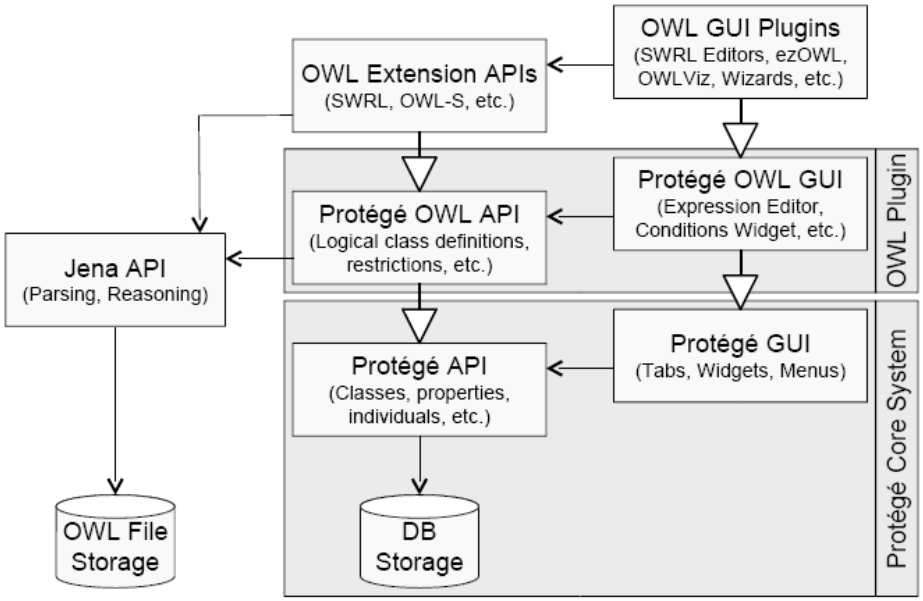


Figure 6.19. Protégé –OWL meta-model [Knublauch et al., 2004]

Pellet is an open-source Java-based DL reasoner developed by Clark & Parsia and it is based on the tableau-algorithm which we already described in chapter 5. Pellet is currently known as the first and only sound and complete DL reasoner that can handle full expressivity OWL- DL and reasoning about enumerated classes (also discussed in this chapter) and qualified cardinality restrictions. Among the major functionalities of Pellet are: checking the consistency of ontologies and checking the entailments, classifying the taxonomy, answering *AB*ox queries and also performing language (or species in terms of OWL-DL, OWL-Lite and OWL-Full) validation and repair.
The architecture of Pellet is given in Figure 6.20 and each module and aspects of reasoning performance optimizations are explained in greater detail in [Sirin et al., 2005]. We used Pellet 1.5.2 for the reasoning process in our BCGO.

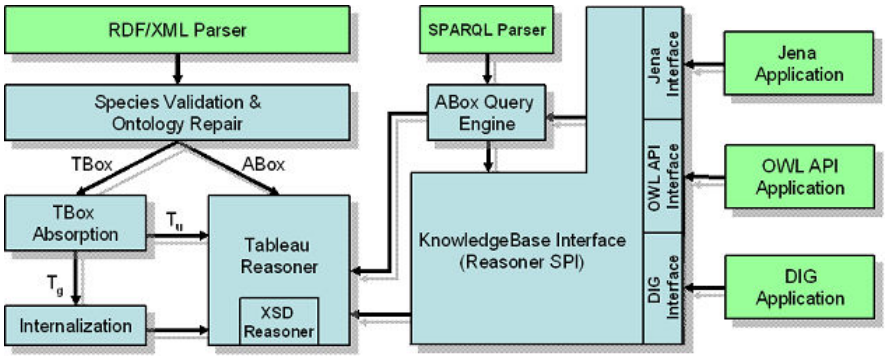


Figure 6.20. Pellet reasoner architecture [Sirin et al., 2005]

6.5. Conclusions

This chapter was devoted to the presentation of the BCGO implementation based on the methodology for the model proposed in chapter 4. The implementation is correspondent to knowledge translation and knowledge refining steps. We addressed the issues of classes in terms of defined versus primitive classes, disjoint classes and subsumption relations. We discussed pitfalls of OWL such as the Open World Reasoning and its connected aspects of universal and existential quantifiers, the unionOf and intersectionOf misunderstanding and misuses.
The properties and instances were described in section 6.2 with regard to types of properties (object and datatype properties), which to use when, the characteristics of object properties (some of them also met in the formal spatial theory) and the problem of domain and range axioms.
The third section of this chapter handled the SWRL rules specific facets, alternative representation of OWL into SWRL. One much more dedicated and specialized way to created alternatives to OWL consists of syntactic sugar rules. We further showed how to combine ontology with rules, by means of DL safe rules. The SWRL rules can further be verified or used in queries over the ontology, an aspect we refer to in chapter 8.
The issue of decidability is directly correlated with DL safety and of high significance for our approach.
The most useful safety conditions are those who boost more expressivity power while keeping the computational power. We gave some example of non-DL and DL-safe rules, the latter leading not to undecidability.
One important note to make is with regard to the spatial properties. The direction in which OWL-DL is moving focuses on representing qualitative spatial information such as translating RCC-8 into OWL-DL (i.e. reflexive properties). This was the argument we also use when adopting RCC-8 for representation of the *CloseTo* spatial relation. Moreover, this aspect comes to emphasize our general approach of qualitative representation.

The implementation of the BCG ontology is a complex one and it is still being refined, as we mentioned that the refining phase is a continually active phase.
Another consideration to be made in connection with the refining phase concerns the level of granularity for our implementation of the breast cancer grading knowledge. We did not represent concepts such as nucleoli, chromatin, or apoptosis (the dead cells that are not to be confused with mitotic figures) or the link between histological grading of breast cancer and other prognostic factors (e.g. hormone receptor status) which belong to a more detail description. Obviously, this kind of information can be included as new knowledge in the ontology when needed, similarly to the information we imported from the NCI/NIH thesaurus to bind and map with breast cancer knowledge.
Acting in this way keeps the ontology alive, in point of fact the life cycle of the ontology modeling and engineering can not go without this phase. To put it in the terms of OWL defined classes, it is necessary and sufficient to work this way such the ontology is consistent and efficiently used in other frameworks (in a reference ontology and/or in our cognitive microscope platform).

Yet, this is a novel and unique BCGO ontology as to our knowledge there is no such attempt neither in the scientific community nor the medical world. Our application ontology has 129 classes, 169 instances, and 86 properties. The total number of restriction is 138, with 79 existential, 49 universal, 1 cardinality restriction, a number of 2 max cardinality and 1 hasValue property.

Having said all that, the main contributions of this chapter are:

- implementation of BCG application ontology model using OWL-DL and SWRL
- detailed presentation of the characteristics, challenges and pitfalls of the formal languages (i.e. how to achieve high expressivity and decidability)

7. Evaluation and Validation of the Model

7.1. Qualitative Evaluation of BCG Ontology

The evaluation of an ontology determines the accuracy and adequacy for its use in a specific context and for a specific goal [Fernandez et al., 2006].
In light or our qualitative approach for the BCGO model, it is obvious that the evaluation of the ontology is to be performed from the same perspective. Let us synthesize the major reasons why the evaluation is given on qualitative grounds:

1. From the very conception of OWL, the purpose of the language itself was to support logical qualitative definition of classes and not quantitative definitions. Cardinality is a logical feature, whilst numerical value of datatype property is not. One could not say "Size of nucleus is 6[cm^2]". Instead, one could define properties as *SmallSize* and *LargeSize* and set value restriction on the class subclasses of Nucleus by *hasValue* restrictions on these. However, the logic reasoner is not able to detect any inconsistency if one has SmallSize set to 6 for a class LargeSize and an instance of this class has value 5 for hasSmallSize property. Despite this view on semantic web languages, some work has been done toward this direction of encapsulating numerical values and it is now possible to capture the quantitative information in OWL as well.
2. One possibility to create definitions with this kind of specification is using SWRL rules, but as mentioned, there is a limitation: these rules have to be DL safe rules in order for the reasoner to take them into consideration and check consistency.
3. Based on our assumption stated in chapter 4, we describe BCG using qualitative expressions and for the inherent quantitative facts coming from the medical knowledge we adopt datatype properties-value restriction approach. In what concerns the spatial concepts, all are qualitative, not quantitative.

While there is no unifying definition for ontology evaluation [Gangemi et al., 2005], some characteristics have been proposed in a taxonomy of evaluation which tackles three major aspects [Brank et al., 2005]:

- vocabulary, hierarchy, semantic relations, syntax
- structure, architecture and design
- content of application

Based on this taxonomy and on existing general tools for evaluating ontologies, [Maiga and Williams, 2009] proposes a tool for evaluation bio-medical ontologies, with three metrics: granularity, scope and ontology integration. We further apply these metrics to our ontology.

1. **Granularity** is viewed as a metric for evaluating relations between levels of ontology hierarchy, set theory, mereology and non-scale dependency (NSD). The NSD granularity considers the primitive relations which have the

key structuring role in ontology, is-a and part-of, with the remark that set theory includes is-a relation and mereology supports part-of relation.

Our BCGO has 129 classes, based on the hierarchy of four main classes, AnatomicalEntity, ConceptualEntity, SpatialEntity and MicroscopicEntity. A number of 169 instances of these classes were generated based on is-a relationship and 86 properties were defined to capture the relationships between instances and between classes. Most of these properties are spatial properties (e.g. SurrBy) described by the spatial representation axioms and theorems and translated into OWL-DL properties or SWRL rules. In terms of levels of hierarchy we have a depth of 10, with an average of 3 siblings and maximum 4 siblings. The total number of restriction is 138, with 79 existential, 49 universal, 1 cardinality restriction, a number of 2 max cardinality and 1 hasValue property.

Additionally, the Open Biological and Biomedical Ontologies foundry (OBO)[6] which is a collaborative ontology development framework also uses metrics such as number of classes, properties, instances, maximum number of siblings, or the depth of the hierarchy to evaluate the ontologies. Although they currently use these primary metrics, OBO is important in terms of principles of ontology development and interoperability (the ability to reuse, share, and integrate data from one or multiple ontologies to another). The OBO effort creates a framework for evaluation and validation of ontologies.

Apart from these metrics, our purpose is to provide a more refined evaluation. In this light, [Tartir et al., 2005] presents several quality metrics categorized as **schema metrics and instance metrics**. The first category *evaluates the ontology design*, while the second analyzes *the degree of reflection* of the real world domain in the ontology. From our point of view, the first category falls on the granularity and the second on the scope metrics. Therefore, we calculate these indicators for our ontology.

RR evaluates the **richness of relationships**, in other words it verifies the connections among classes in terms of how much of the connections are rich relationships compared to all relationships (which include rich relationships and inheritance relationships). The formula we adopt is:

$$RR = \frac{\sum P}{\sum P + \sum SC} = \frac{86}{255} = 0.33$$

where P stands for the relationships among classes, SC for the number of inheritance relationships which is equivalent to the number of subclasses.
In our case, RR is 0.33 which indicates that most of the relations from our ontology are class-subclass (is-a) relationships, since RR it's more close to 0 than to 1. When an ontology has mostly class-subclass relation than others relations, it means it is not highly richer than a taxonomy.

[6] http://www.obofoundry.org/, last accessed July 2010

Another metric is the **Attribute Richness** (**AR**), which articulates the number of all attributes (properties) divided by the number of all classes from the ontology. This metric is very important for it reveals the amount of information used in instance data, on one hand and quality of the ontology design, on the other hand. We get a value of:

$$AR = \frac{\sum P}{\sum C} = \frac{86}{129} = 0.66$$

which shows that lot of knowledge is conveyed in the ontology (close to 1). We mention that all set theory and mereology relations we used were included in the computation of number of relationships where the formulas required them.

2. Scope or the degree of reflection is designated by instance metrics. They classify into knowledge-base metrics and class richness.

Average Population (**AP**) is one of the knowledge-base metrics, which shows how many instances were defined in the knowledge-base K divided by the number of classes in the ontology. The result shows a very good population of the K.

$$AP = \frac{\sum Inst}{\sum C} = \frac{169}{129} = 1.31 \cong 1$$

If we apply the **Class Richness** (**CR**) metric which indicates the average number or classes from the K that have instances, divided by all classes from the ontology, we obtain:

$$CR = \frac{\sum C'}{\sum C} = \frac{147}{129} = 1.13 \cong 1$$

The result shows the knowledge-base $\mathcal{K}$ has almost all data that exemplifies all knowledge from the ontology. Note that our ontology is an application ontology not a reference ontology and the number of instances are directly influenced by the complexity of the domain of knowledge we represent. Further more, the development of the ontology is still ongoing.

3. **Integration** metric analyzes the level of connectedness with concepts or relations from related ontology. Breast Cancer grading ontology is a lightweight or application ontology and to the best of our knowledge, no other ontology representing the same domain has been developed. In our design framework, in the knowledge acquisition phase, we pertain to NGS and NCI/NIH thesaurus.
A subset of NCI thesaurus' breast anatomy and breast cancer related concepts was imported to our BCG application ontology (e.g. concepts such as Disease, Patient). This subset together with the NGS terminology contributed to building the explicit knowledge model.
Formally, we define the **Ontology Integration** (**OI**) metric as:

$$OI = \frac{\sum IC}{\sum C} = \frac{35}{129} = 0.27$$

which indicates that the level of connectedness with respect to classes is low, yet expected, considering all the facts mentioned above.
Since we integrated a subset from a related thesaurus, we envision a vice versa action - an integration of our ontology into a reference ontology, which is of future interest for our work.

Evaluating from the reasoning perspective, the tableau algorithm has to meet three requirements: soundness, completeness and termination.
The result of logical consistent ontology was obtained within 2*:*48 seconds. The time refers to the termination requirement of the tableau-based algorithm. We obtained specific results in a finite time (2:48 sec) therefore the termination requirement was met (the specific results deal with the logic consistency- the soundness and completeness requirements which we discuss in section 5.3.1. When the verification is complete, we achieved a complete model.
Once this task is accomplished, the reasoner generates the inferred model from the explicit model, through a complete classification of hierarchy. Both models can be visualized and compared using OWLViz, Jambalaya or other visualization tools.

7.2. Syntactic Constraints of OWL- DL versus OWL- Full

Due to the nature of semantic web language, one aspect needs particular care in evaluation of the BCGO model. This regards the syntactic constraints of the OWL-DL. In order to satisfy the decidability task so that to make BCGO consistent, several syntactic constraints are imposed over the language.

1. The first constraint deals with the **cardinality restriction on transitive properties**. If there are such restrictions, the expressivity of the language is OWL-Full instead of OWL-DL.
The cardinality restrictions specify that a class of individuals may have at least (minCardinality), at most (max Cardinality) or exactly a specified number or relationships with other individuals or datatype values of properties.
For instance, we say that class "*Slide hasNottinghamGrading* some *NottinghamGrading"* and that "*Slide hasNottinghamGrading* max 1". A maximum cardinality restriction is set for the individuals from *Slide* class such that they may participate in only one relationship with individuals from class *NottinghamGrading*. In other words, any *Slide* can have only one *NottinghamGrading*. This restriction does not affect the expressivity of the language, (in terms of this specific condition) because the *hasNottinghamGrading* is not a transitive property.
The property *isLocatedIn* is a transitive property. An example that would fail the test of OWL-DL with this property is the following (Figure 7.1). According to the definition, *Mitosis* is any *NucleusDivisio*n that is not located in *Tubule*, but is located in I*nvasiveFrame*, where the *isLocatedIn* has a cardinality restriction set to maximum one relationship along it.

OWL:
Class (Mitosis complete NucleusDivision
complementOf (restriction (isLocatedIn someValuesFrom Tubule))
complementOf (restriction (isLocatedIn allValuesFrom Tubule))
restriction (isLocatedIn someValuesFrom InvasiveFrame)
restriction (isLocatedIn max 1))

OWL-DL:
$Mitosis \equiv NuclearDivision \sqcap$
$\neg\exists isLocatedIn.Tubule \sqcap$
$\neg\forall isLocatedIn.Tubule \sqcap$
$\exists isLocatedIn.InvasiveFrame \sqcap$
$\leq isLocatedIn$ max 1

Figure 7.1. OWL-DL constraints of cardinality restriction on transitive property

Otherwise stated, a *Mitosis* has *isLocatedIn* transitive relationships with individuals from two different classes, which is not possible due to the restriction on the numbers of relationships.

2. Also related to transitive properties, a fact that we already mentioned is **that transitive properties *cannot be functional.*** The example of *hasIdentifier* restriction for a *BreastCancerPatient* is illustrative; a breast cancer patient cannot be connected via *hasIdentifier* relationship with two different individuals from the *Identifier* class.

3. **No classes or properties in enumeration.** Enumeration may contain only individuals that belong to a class for an OWL-DL representation to be satisfied. In OWL-Full, a class may be an individual and an individual a class as well.
In this example, the Slide class contains an enumeration of individuals (Figure 7.2). The enumeration is not complete, there could be more Slide individuals, but for the sake of simplicity we added two. In this case, the OWL-DL test holds, as the enumeration does not contain any properties or classes.

OWL:
Class (Slide complete VirtualSpecimen
restriction (hasIdentifier someValuesFrom Identifier)
restriction (hasIdentifier allValuesFrom Identifier)
restriction (hasNottinghamGrading someValuesFrom NottinghamGrading)
restriction (hasNottinghamGrading allValuesFrom NottinghamGrading)
restriction (enumeration {Slide1 Slide2}))

Figure 7.2. Enumeration OWL-DL restriction

4. **No super- or sub-properties of annotation properties.** The annotation properties allow all components of the ontologies (classes, properties, individuals) including the ontology itself to be annotated using meta-data. Whilst OWL-Full does not put any constraint on annotation properties, in OWL-DL there are two major constraints: the filler for the annotation property must by a datatype value, an URI (Unique Resource Identifier) or an individual, and the second one, there must be no hierarchy of properties for an annotation property, unlike the case for object or datatype properties.
There are several pre-defined OWL annotation properties which can be used to annotate the property (e.g. owl: versionInfo, rdfs:label, rdfs:comment).

5. **No properties with class as range**. Since we just discussed about the annotation properties, a situation which leads to inconsistency in OWL-DL terms, is for some Protégé framework pre-defined annotation properties. For instance in the case of protégé:AllowedParent or protégé: todoProperty , if a class (e.g. Datatype property for protégé: todoProperty) is set up as range.
In order to check for other inconsistencies as such, one possibility is to give the reasoner a query to find all instances that are subclasses of rdfs: Class and are also the rdfs: range of some other owl: Class. For instance, if for the *isLocatedIn* property we set up the range to Ducts, this would classify several classes as inconsistent, because there are other individuals linked via this property with individuals from classes other than Ducts. Instead if hasSize property has range Size, this would lead to consistency since all individuals engaged in a relationship with individuals from class Size can not have this relationship with other individuals than Size individuals.

6. **No subclasses of RDF classes and no metaclasses**. Since OWL is an extension of RDF or more precisely of RDF Semantics, any graph forms a OWL-Full ontology, and OWL-Full ontologies can include random RDF content, whilst the OWL-DL puts the restriction on the hierarchy of RDF classes. In OWL Full, like previously mentioned, a class can be treated as an instance of meta class, using the owl:sameAs construct to define equality where a meta-class is a Class that is a subClassOf owl:Class. In other words, a metaclass is a class whose instances are themselves classes. In our ontology there are neither metaclasses nor subclasses of RDF classes.

7. **No import of system ontologies**. Although OWL supports the reuse and sharing of ontologies by means of import features, the OWL-DL does not allow any import of ontologies. The reason is that the description of classes, instances and properties from the ontology one imports may not be following the OWL-DL restriction. However, if there are imports, such as in our case- redirected imports (e.g. swrl.owl, tbox.owl, sqwrl.owl), the OWL sublanguage may be OWL-Full.

7.3. BCG Ontology Validation

The representation of a domain in ontology relies on a wise strategy to interact with and interconnect the elements of the semantic web: standards of languages and reasoning systems, knowledge-bases, ontologies and inference engines. As in

medical domain, reaching the consensus with respect to knowledge representation is the first step to success for a later integration. The purpose of a knowledge representation system goes beyond storing concept definitions and assertions. It requires continuously validation of their logic, consistency and quality. The validation process of ontology is done on several levels, in a spiral-shape. At the beginning it deals with validation of the modeling phase, followed, after satisfactory results, by the development phase. It keeps on refining until the ontology can be integrated into a larger ontology and maintained there, since all ontologies evolve. The process advances to a more complex layer of the larger ontology validation.

In the Protégé framework we used for building the BCGO ontology, there are two different techniques which complement each other for visualizing the ontology (Table 7.1).

OWLViz	**Jambalaya**
• based on GraphViz DOT technique	• uses Simple Hierarchical Multi-Perspective(SHriMP) technique
• class hierarchy visualization • incremental navigation of asserted & inferred model • *is-a* relationship oriented	• visualization, exploration and understanding of knowledge bases • *hasInstance* and *hasSubclass* relationships oriented
• plug-in tool in Protégé	• plug-in tool in Protégé

Table 7.1. Visualization techniques

The OWLViz tool allows visualization of the asserted and inferred class hierarchies as graphs. Unless the ontology that is being edited has been classified at least once, the classes in the inferred hierarchy will be displayed without any edges connecting them – this is because the edges represent *inferred subsumption relationships*, which do not exist until classification has taken place. Otherwise, the edges show *is-a* relationship from the ontology.
A fragment of the inferred model is given in Figure 7.3 [Tutac et al., 2009d].

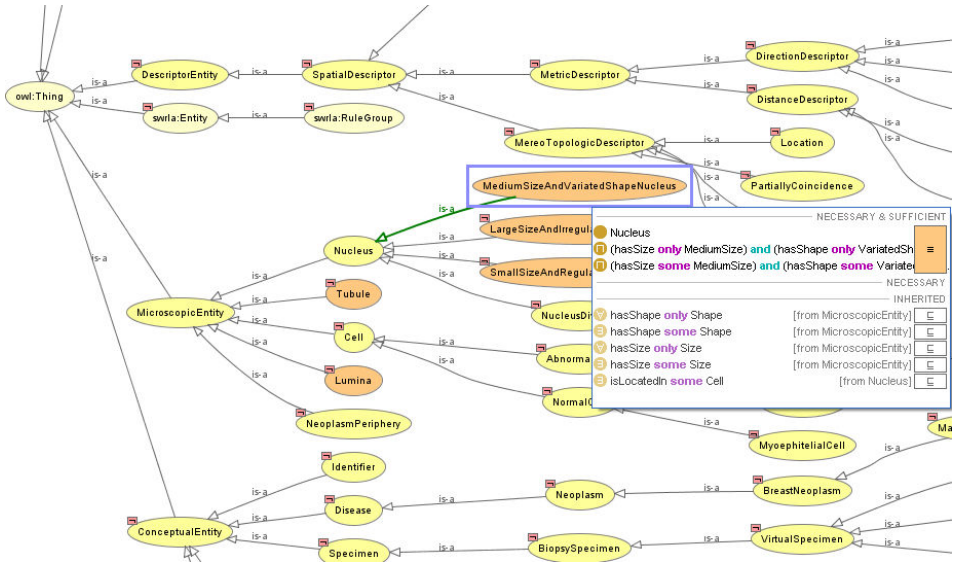


Figure 7.3. Fragment of the inferred model of ontology– OWLViz view

Due to the large size of the ontology, the Jambalaya graphic displays only a small fragment of it in Figure 7.4. Note that the *hasInstance* relations are marked with red, whilst *is-a* relations are represented in blue. For the instances of classes a rhomb symbol is used and a circle symbol is assigned to classes.

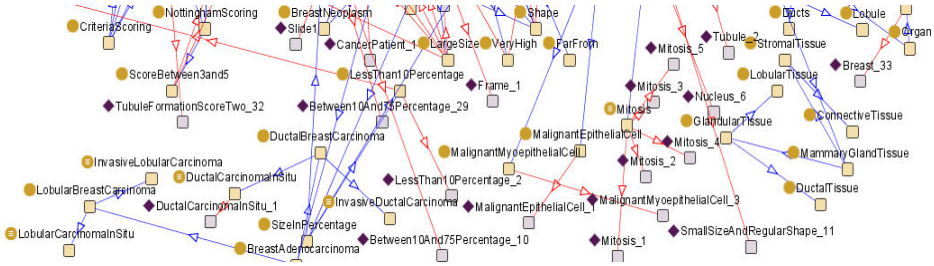


Figure 7.4. Jambalaya View: *classes* are marked with circle, *instances* are marked with rhomb, *has Instance* relations are marked with red, *is-a* relations with blue

7.3.1. Semantic Retrieval

We make a distinction between the types of queries we can perform, in terms of what kind of knowledge representation we have. Querying the ontology with SQWRL does not consist of a database query although it is based on DataLog. It consists of a knowledge-base query, which we call a semantic query.

Amongst the semantic query modalities, we distinguish four types of queries: RDF/SPARQL, SQWRL, Jambalaya and Pellet queries (Table 7.2).

Types of query	Description & Languages
Database query	information is structured in a *database* e.g. DataLog, SQL
Visual query	information is represented by the *image* content (image descriptors) e.g. QBIC (CBIR)
Semantic query	information is structured in a *knowledge-base*, performs reasoning e.g. SPARQL, SQWRL

Table 7.2. Types of query

For the RDF-like query, we first recall the RDF representation in terms of a triple containing subject, object and the relationship between subject and object (Figure 7.5).

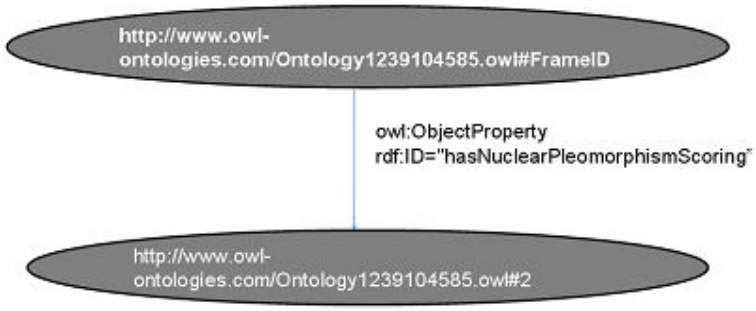


Figure 7.5. RDF representation

The RDF query given in Figure 7.6 asks the system to find all frames which have Nuclear Pleomorphism score 2. This kind of query can be introduced by the computer scientist as well as by the pathologist [Tutac et al., 2009d].

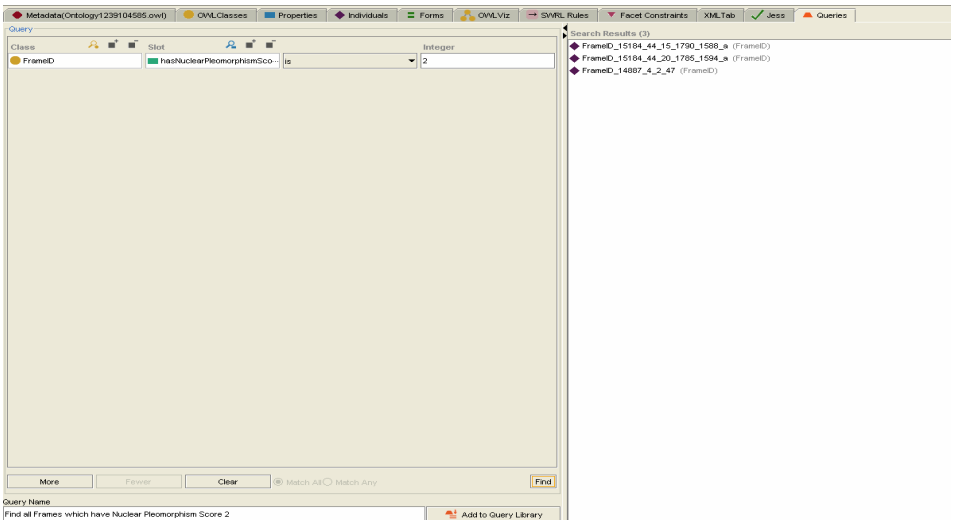


Figure 7.6. RDF query for all *Frames* having *NucleaPleomorphimScore2*

Considering the following query:
"Find all Slides which have NuclearPleomorphism Score 2 and Mitotic Count 3"

and given a SQWRL (Semantic Query Web Rule Language) formulation as the following:

$$\text{Slide}(?x) \wedge \text{hasNuclearPleomorphismScoring}(?x,?\text{nuclearscore}) \wedge \text{hasMitoticScore}(?x,?\text{mitoticcore}) \rightarrow \text{sqwr:select}(?x,?\text{nuclearscore},?\text{mitoticscore})$$

the problem is that this query can not be computable since the DL constraints (discussed in chapter 2) are not followed. We can not translate this rule in a DL safe rule, as there are two different variables (?nuclearscore, ?mitoticscore?) shared between the antecedent part of the rule (the body of the rule) and the consequent (the head of the rule). It should be noted that SQWRL provides an SQL-like operation to retrieve knowledge from an OWL ontology, using SWRL syntax and semantics. In other words SWQRL is a query language dedicated to OWL representations, rather than a pure SQL database query. The advantages of SQWRL for OWL representation over other query languages (e.g. SPARQL) are: conciseness, readability and semantically robustness. These major points are detailed in [O'Connor and Das, 2009].

The next formulation gives a correct answer and satisfies the DL safety conditions.

$$\text{Slide}(?x) \wedge \text{hasNuclearPleomorphismScoring}(?x,?\text{nuclearscore}) \wedge \text{NuclearPleomorphismScoring}(?\text{nuclearscore}, 2) \wedge \text{hasMitoticCount}(?x,?\text{mitoticcore}) \wedge \text{MitoticCount}(?\text{mitoticscore}, 3) \rightarrow \text{sqwr:select}(?x)$$

In order to run a SQWRL type of query, let us consider the mitosis example. If we want to retrieve all mitosis instances from the ontology, the part related to the eccentricity property from definition of *Nucleus* and *Mitosis* is beforehand necessary for the rule engine. In this particular case, the eccentricity property makes the distinction between nuclei and mitosis (Figure 7.7). The rules are automatically processed and activated by the computer scientist. It is however more difficult for a non-trained pathologist to use this type of query than the RDF query, unless there is a high level Graphical User Interface (GUI) implemented to help him in doing so. The results of the query are very important for the prognosis traceability (Figure 7.8).

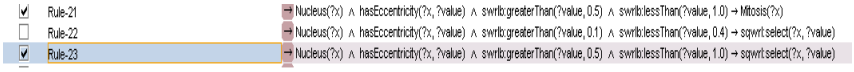


Figure 7.7. SQWRL: Show all nuclei that are mitosis. Rule 21 and 23 are selected and activated in order to process the mitosis definition and to further retrieve all mitosis instances from the ontology

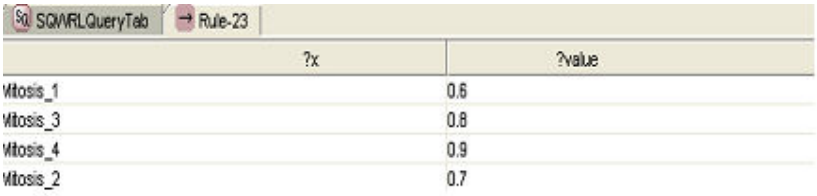


Figure 7.8. Query results for Mitosis with the corresponding eccentricity value in accordance with rule 23

A Jambalaya query is given in terms of hierarchical representation as there are various ways to display the ontology. Another way of giving Jambalaya query is by user input: for instance, to find all concepts that contain the expression DuctalBreastCarcinoma.

At this point, one signification question arises: what about an evaluation in terms of precision and recall since we are discussing about retrieval? While precision and recall results have been computed for our methods of automated grading based on image processing algorithms ([Dalle et al., 2008], [Dalle et al., 2009], [Veillard et al., 2010], [Huang et a;l, 2010]), and some results for the image-driven approach from the CBIR-CBR combined framework ([Tutac at al., 2009a], chapter 8), a computation of precision and recall for the ontology-driven is very difficult for multiple reasons.
Firstly, we do not apply a classic CBIR in our method to compare a query image to the images from the database, according to the signature of each image. This is due to the fact that the semantic annotation of images is based on the BCGO description, not on the features extracted from the images. The ontology drives the process of annotation of the images by connecting the semantic level with the image level. In terms of language programming, this is showed in Table 4.3. Secondly, this

would have been difficult as there are images with different magnification (for the criterion of the BCG) or there are types of structures such as DCIS. Thirdly, in order to obtain relevant results, the pathologists would need to provide us, in the first place, with more medical cases than we have so far or to annotate all frames and grade all frames and slides. This is a very consuming and tedious task (we discuss about this issue in chapter 8 and 9 as well). Lastly, besides BCGO indexing our approach applies an explorative philosophy. Therefore the whole demarche is completely different and is very difficult to quantitatively evaluate it since there is no benchmark to compare with. In chapter 8 we present MICO as a solution to these problems, as a first automated benchmark for evaluation of breast cancer biopsies.

7.3.2. Medical & OBO Validation

Medical reasoning deals with the clarity of description from the medical standpoint. It also checks whether there are some concepts or descriptions of rules missing. This requires patience and running several automated test before one could be able to see if something is not working properly.
Based on our collaboration with the National University Hospital (NUH) from Singapore, the Pathology Department provided us the materials to study on 20 breast cancer slides/patient which corresponds to 80.000 hyper power-fields (high-resolution frames, where a slide is composed of 4000 hyper-fields) of 1024*1024 size acquired at 40X magnification and stained with H&E marker. These patient cases, together with the information extracted from the medical reports related to NGS make the ground-truth for our medical validation. The NCI thesaurus completes the medical knowledge with breast cancer related information. As the ontology is continuously developing, the medical feedback is also an interactive process. It can be argued however, that a practical impact in a clinical setting would emphasize more the efficiency of our approach. This fact is a mater of concern for our future works.
Looking from the computer science point of view, one note of high importance is that, when starting to develop an ontology, some structures of the definitions of concepts may look differently compared to the medical view. Hence, the consensus agreement plays a key role through the whole life cycle of the BCGO.
In addition, we integrated the ontology into BioPortal, as part of the Open Biomedical Ontologies (OBO) foundry. As discussed in section 7.1, the OBO uses some primary metrics to evaluate ontologies. The advantages of this integration regard the mapping and inter-operability aspects. As an application ontology, BCGO could be further integrated into upper-ontologies, or concepts from it could be mapped with similar concepts from other ontologies.

Our ontology can be freely accessed at IPAL website[7] and it has been recently involved in the first contacts between IPAL lab and AGFA Healthcare company[8] through our TRIBVN[9] collaboration.

[7] http://ipal.i2r.a-star.edu.sg/project_MICO.htm, last accessed July 2010
[8] http://www.agfa.com/en/he/home.jsp, last accessed July 2010
[9] http://www.tribvn.com/page.php, last accessed July 2010

7.4. Concluding Remarks

In this chapter we addressed the issue of evaluation and validation of an ontology. Since in our approach (which will be integrated into the MICO presented in the next chapter), the ontology plays a key role in the image processing and in providing a second opinion of grading, an evaluation of our ontology is required. Given the reasons for a qualitative evaluation, we rely on the metrics proposed by [Tartir et al., 2005] in light of [Maiga and Williams, 2009] direction, as there is no unifying methodology for such task. One important thing we want to stress is that the evaluation is strongly dependent on the objectives of an ontology and what domain addresses, well reflected in the metrics of granularity, scope and integration. Based on these analyses, we could say that our lightweight ontology representing the domain of breast cancer grading and more specifically the Nottingham Grading System has been well designed, the degree of domain reflection is good and it opens promising perspectives for inter-operability and mapping with upper-ontologies.

Another note of high significance regarding the syntactic constraints of OWL-DL versus OWL-Full is that in an ontology there could be fragments of OWL-DL and OWL-Full. This does not mean that the reasoner cannot classify and check consistency. The classic argument from DL to be decidable is not valid in the sense that you are never sure about the validity of the answer because of the inherent incompleteness of the model. Moreover, OWL Full, just like FOL is semi-decidable and will give you a decision procedure if the answer is known to be true or false. Hence, if one of the purposes the curator of an ontology has is to share this ontology to be used in other related ontologies, this single constraint may not favor the OWL-DL, bearing in mind that one can still obtain decidability.

The validation of the ontology is approached from several perspectives. The semantic validation is carried by means of different types of semantic queries and we show how to distinguish among them in our examples. The medical validation concerns the feedback received from the medical experts we are collaborating with as part of the refining phase from the theoretical model proposed in chapter 4. Further evaluations on the applicability of our approach in a clinical setting in terms of quality enhancement and time reducing in the work of the pathologists, are however needed. The OBO foundry is a repository for ontologies developed in medical field which can be reused in other ontologies. The BCG ontology is the only ontology that deals with representation of the breast cancer grading domain. From this point of view an evaluation of BCGO in comparison with another ontology is not relevant.

Synthesizing, the key contributions of this chapter are:

- evaluation of the BCG ontology using qualitative paradigm
- ontology consistency and decidability in terms of OWL-DL
- validation of the BCGO using semantic queries, medical feedback and reusability and sharing criteria in ontologies community

8. Model Applicability. MICO- Cognitive Virtual Microscope Prototype

Among the various contexts in which semantics, ontological- knowledge representation can be used and exploit, in the IPAL[10] team we advance a novel setting: a cognitive virtual microscopy platform for breast cancer histopathology.
The core reason we choose this context is that the field of microscopy is experiencing a new revolution, exactly in the same way as satellite imaging systems do, mainly due to dramatically improved storage and acquisition capabilities. The urge for intelligent interaction for image exploration tasks is gaining momentum.
Secondly in the medical context as mentioned in chapter 4, histopathology became widely accepted as a powerful "gold standard" for prognosis in mainstream diseases such as breast cancer allowing to narrow borderline diagnosis issued from classical macroscopic analysis (i.e. mammography and ultrasonography). Yet, one of the major issues in the histopathological image analysis process is the subjectivity, inconsistency and tediousness of human manual work. This directly impacts diagnosis and treatment decisions. The importance of such direction is proven by the fact that MICO has been recently granted by the Technologies for Health Program (TecSan) of the French National Research Agency[11] (Agence Nationale de la Recherche- ANR), as a project which involves partners from academic, medical and industry fields.

This chapter gives a brief survey on the technological roots of cognitive virtual microscopy in the first section. Then, it introduces the Cognitive Microscope (MICO) platform built by IPAL team and addresses the characteristics of a cognitive system which are to be found in this system as well. It finally concludes with some contributions.

8.1. Virtual Microscopy

A virtual microscopy system is generally seen as an emulation of real microscopes, together with tools enabling image analysis and storing.
It has been reported that virtual-slide telepathology (consultation, education and research using telecommunications to transmit data and images between two or more site remotely located from each other) is a growing industry that will greatly expand over the next ten years [Weinstein, 2007], [Mulrane et al., 2008]. Partly due to the amount of scanned data for just a patient case, major technological hurdles are still to be overcome.
However, although network transmission or storage issues are not yet settled, Whole Slide Imaging (WSI) systems commonly use client-server architecture and most of them are able to work either as standalone or as internet connected

[10] http://ipal.i2r.a-star.edu.sg/project_MICO.htm, last accessed July 2010
[11] http://www.agence-nationale-recherche.fr/, last accessed October 2010

systems. To facilitate access to any image sub-region at the required resolution, the system providers use a tiled organization, stored in a pyramidal structure wherein each level contains pre-computed images at lower resolutions (e.g. 1/4th, 1/16th) [Soenksen, 2005]. Some platforms are dedicated to a specific WSI format (generally the one provided by the slide scanner providers) while other platforms are able to handle various formats (tiled format or JPEG2000 format).

A comparison of 31 digital slide commercial systems was given in a recent survey [Rojo et al., 2006]. With such systems, pathologists can browse the images as if they were using a real microscope (multi-scale virtual browsing) and perform "slide conferencing", through synchronized cooperative sessions, allowing collaborative image annotation. A few systems also provide image analysis tools (segmentation, classification) for detecting regions of interest, cells, nucleus or membrane.
There are also a few free digital slide systems designed mainly for educational purposes. They allow building microscopic image libraries, providing a tool to navigate through a WSI and for manual annotations. The most impressive of these tools is WebMicroscope [Lundin et al., 2004]. Images of the scanned slides are stored in a database on a server connected to the Internet. On the client side, a plug-in — installed on standard web browser — allows the user to query the server for a slide, to navigate through it and to annotate it.
One of the most advanced research team in virtual microscopy established the Rapid Breast Care Service to provide a woman who has a positive digital mammography study with the results of her laboratory analysis the same day [Weinstein et al., 2005]. DMetrix Company in connection with this research team has developed a digital-image archiving system called the "Arizona" system [Weinstein et al., 2007]. Users are able to access image data and metadata through a secure Web site. However, this highly efficient system does not use expertise modeling for emulating knowledge enhanced virtual microscope. Moreover, these systems did not intend to support medical decision-making.
To sum up, none of these systems is designed as an integrated computer-aided prognosis platform. In addition, they lack operational methods for very large scale image analysis and effective knowledge management capacities.
Cognitive virtual microscopy emerged from virtual microscopy inspired by research challenges in cognitive vision applied to many fields, but in our context to medical imaging. [Shanmugaratnam, 2007] reported that errors and discrepancies in histopathology are not uncommon; various studies estimated that 5% of all reports contain medical errors depending on the methods used for error detection, and the definition of what counts as an error. Some of these errors are cognitive, or induced by a lack of cognitive attributes. Thus, a cognitive microscope could help decreasing these cognitive errors as well as over-grading cases.
On another hand, new microscopic devices dramatically need smart exploration strategies in the same way as robots need it when exploring hostile environments. In our context, the histopathological image can actually be considered as a challenging research playground of a new kind of autonomous, intelligent robot that is the virtual microscope system.
But, from an operational point of view, how can we define the cognitive capabilities of a cognitive virtual microscope?

8.2. Cognitive Virtual Microscopy

To answer this fundamental question, which is challenging even in the field of classical robotic systems, we refer ourselves to the inspiring report of the European Research Network for Cognitive Computer Vision Systems [Auer et al., 2005], [Loménie et al., 2009].
Stating that computer vision is still a brittle technology, this group dramatically appeals for extending the performances of state-of-the-art computer vision systems in terms of:

- robustness: the state (or capacity) of a system that can **absorb** a stressor **without adapting;**
- resilience: the state (or capacity) of a system that is **able to adapt** to a stressor;
- **adaptive capacities:** more operational than resilience being related to re-configurability by adding cognitive capabilities like:
- **learning** and Adapting, entailing a subtle difference between the two concepts;
- weighting alternative solutions;
- developing new strategies to analyze and interpret;
- **generalizing to new applicative contexts or domains;**
- **communicating with others systems, including humans.**

A description of the nature of a cognitive system into three axes (scientific foundations, functional capabilities and instantiated competencies) is summed in Table 8.1. In bold police are the items in which a cognitive microscope as ours can be involved at short/mid-term.

Scientific foundations	**Functional capabilities**	**Instantiated competencies**
Visual sensing	**Detection& localization**	▪ Cognitive home assistant ▪ **Advanced driver assistance system** ▪ Cognitive assessment of behavior of shoppers in retail outlets ▪ Monitoring of adaptive advertisements ▪ Interactive toys assistance for elderly and infirm
Architecture	Tracking	
Representation	Classification	
Memory	prediction	
Learning	**Concept formation & visualization**	
Recognition	**Inter-agent communication& expression**	
Deliberation& Reasoning	Visual-motor coordination	
Planning	Embodied exploration	
Communication		
Action		

Table 8.1. Nature of cognitive systems

As stated previously, cognitive vision is now at its pre-paradigmatic status and a few challenges were identified as primordial in the coming years among which:

- methods of continuous learning;
- utilization and advancement of systems engineering methodologies;
- **development of complete systems with well defined competences;**
- **research tools for the community like software platforms or generic modules;**

These challenges assume two underlying outcomes of this research lines: to leverage the capabilities of designing robust vision software products, **to design adaptable and adaptive interfaces between end-users and computer systems.**
In that perspective, our virtual microscope exploring and eventually acting on microscopic images constitutes a good paradigm as a device to perform cognitive tasks in a closed universe [Loménie et al., 2009].
The ongoing ambition is to develop a complete system for microscopic image exploration with added values like continuous learning and eventually knowledge discovery capabilities in the long term.

8.3. MICO – Cognitive Microscope prototype

The cognitive microscope aims at enhancing the diagnosis process through a synergy between knowledge, context, cognition and experience based on a user-centered approach to provide visual prognosis assistance to pathologists [Roux et al., 2009b], [Loménie et al., 2009], [Racoceanu et al., 2009].
As for now, from the cognitive point of view as defined in the previous section the system is able to:

- detect and localize histological objects of interest;
- visualize histological concepts by a seamless mapping between spatial representation given by the image analysis algorithms and symbolical concepts via knowledge representation tools;
- enhance the cognitive cooperation between the human agent and the virtual computerized system; MICO is intended to be used in the field of pathology research as a Virtual Consultant

The global framework of the cognitive virtual microscope MICO is represented in Figure 8.1. It is built on a modular design, enabling a flexible configuration and adaptation of the system to different applications in the area of histopathology, cytohematology and bacteriology/virology. The platform is divided into several modules communicating via a middleware based on semantic web technology. This distributed structure supports easy development, incremental evolution, self adaptation and remote applications such as telepathology and slide conferencing. We present the modules of MICO, showing the integration of our ontological approach (knowledge-guidance/ontology) in the general framework.

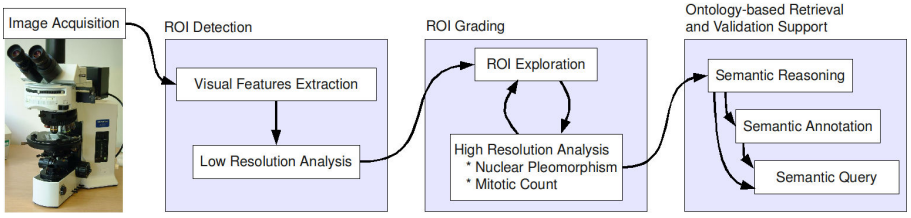


Figure 8.1.Functional framework of the cognitive virtual microscope MICO

Image acquisition. We use a standard optical microscope which can be found in most of the analysis laboratories in pathology or bacteriology (in our case, an optical microscope Olympus BX51, with 4X/10X/40X/60X/100X possible magnifications, Prior "H101A ProScanII" motorized X/Y stage and Z focus with a travel range of 114mm×75mm and a minimum step size of 0.02μm, and a 1280×1024 pixels digital camera MediaCybernetics "EvolutionLC color" IEEE1394 MegaPixel). We use a Media-Cybernetics controller connected to the microscope to perform **an acquisition** of high power fields/frames (in our study at 40 X magnifications according to the request of the pathologist for the high resolution grading algorithms) from a slide put under the microscope. After acquisition, the frames are converted to DICOM format and stored into an open source PACS (Picture Archiving and Communication System) system (dcm4chee1) according to the hospital standards. The acquired 40X high power fields are stitched in order to obtain the WSI. Figure 8.2 shows a WSI example from the MICO platform.

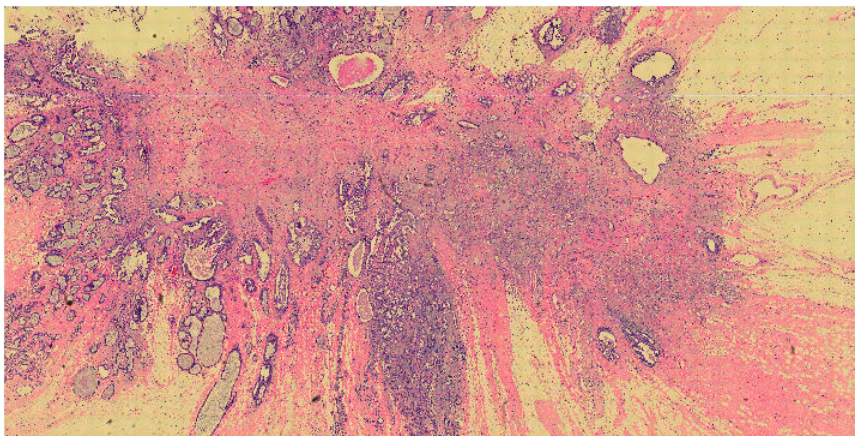


Figure 8.2. Example of WSI from MICO platform

Region of interest (ROI) detection is a fundamental phase of breast cancer grading for histopathological images. Pathologists identify ROIs to efficiently select the most important invasive areas for the grading process. Since neither pathologists nor computers are able to explore every detail at high magnification within a reasonable time, the effective and efficient choice of the ROI is thus a critical step. In this study, ROI detection is denoted as a classification problem. The

low magnification analysis will determine if a given region is an invasive area in a similar manner a pathologist would do when evaluating a biopsy. In order to mimic this behavior, the IPAL team exploits the relationship between human vision and neuroscience [Huang et al., 2010].
Figure 8.3 depicts the process of ROI construction. The visual signal of the human visual system can be described using color pairs (red-green and blue-yellow) and luminance information. In our method, our feature extraction algorithms include intensity, color and texture perception. The resolution of the image we work with in this module is 8.3 μm/pixel at 1.2X (12 times) magnification. However, the details of the features extraction or classification methods are not falling into our primarily scope of formal knowledge integration into this cognitive virtual microscope system.

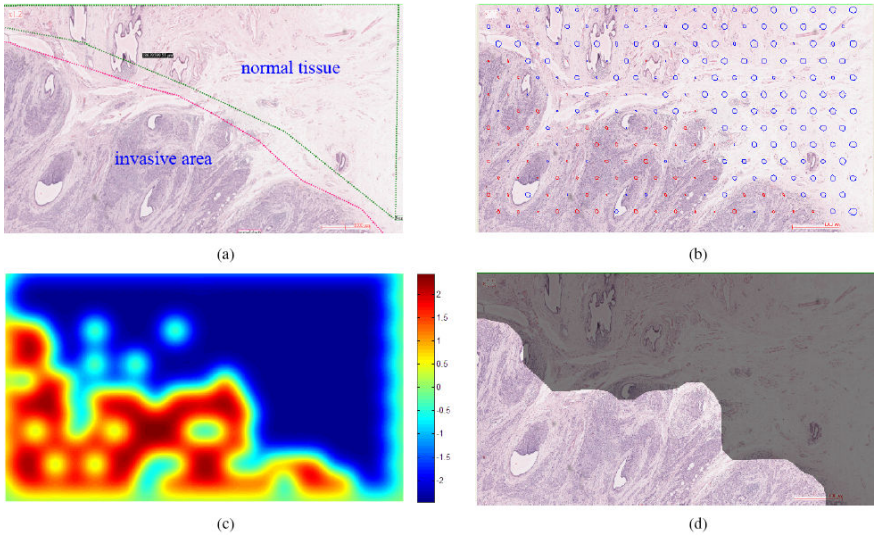


Figure 8.3. An example of ROI construction. 8(a) Ground truth provided by pathologists on the input image with two regions: invasive and not invasive areas. 8(b) Result of feature extraction and classification methods on a set of equally distributed testing points. The results are illustrated as circles. The red-circles indicate positive areas, and the blue-circles are negative areas. 8(c) Low-pass filtering in order to estimate the characteristics on the areas between the testing points. 8(d) The region of interest is obtained by thresholding [Huang et al., 2010]

Multi-scale global nuclear pleomorphism score. Once the ROI is detected, a map of local cancer grades is established using a multi-scale, computational geometry-based dynamic sampling method [Veillard et al., 2010] combined with high resolution image analysis techniques (40X magnification). This approach is used for the nuclear pleomorphism scoring.
We rely on segmentation of only critical cell nuclei affecting the score based on the conclusions of a previous approach [Dalle et al., 2008]. The detection of the cell nuclei includes three distinctive stages: detection of the cell nuclei in which color, shape and size features are extracted, followed by segmentation of nucleus boundaries using polar space and post-processing with snake-algorithm (Figure 8.4), and finally scoring of a population of cell nuclei.

The global score is computed from the frames having the highest Nottingham grade extracted by our high resolution nuclear pleomorphism grading algorithm applied to the entire ROI. However, a single ROI can be potentially very large, up to several thousands of high resolution frames in some cases, making impractical such exhaustive analysis.
Meanwhile, the current research effort in histopathology image analysis focuses essentially on processing high resolution frames and does not consider the problem at the WSI level. Therefore, we filled the gap by developing an innovative method to rapidly identify the frames of interest necessary for a global grading.

Our algorithm aims to establish a map of the nuclear pleomorphism levels encountered within the individual ROIs. In practice, the grading is performed by the pathologist based on a few frames of interest selected for showing the highest grade of cancer in the slide. Usually, the most cancerous areas can be identified with the regions having the highest degree of nuclear pleomorphism. It is the best indicator to obtain a global map of the cancer because it can be assessed locally (frame wise) and is precisely quantifiable in a wide, continuous range (although the Nottingham grading system discretizes it). In comparison, mitoses are too sparse to be a good local indicator and tubular formations do not distinguish between different advanced cancers where tubules are absent.

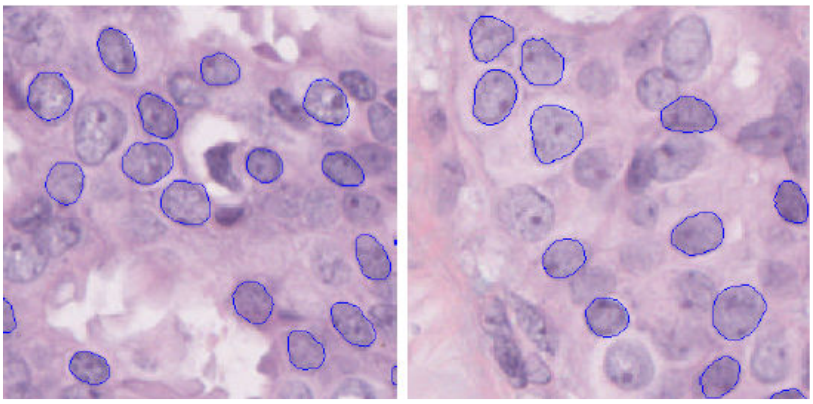


Figure 8.4. Cell nuclei segmentation based on the snake algorithm [Dalle et al., 2009]

Figure 8.5 shows the highest grading area obtained from the resulting map. Global grading is performed using the frames selected from this area. The results of these methods are also discussed in [Dalle et al., 2009], [Veillard et al., 2010], [Huang et al., 2010].

Ontology-driven mitosis and tubule formation scoring. When features extraction algorithms based on raw images provide powerful support for object segmentation and classification, an ontological representation to guide the algorithms is not absolutely necessary. This is the case of nuclear pleomorphism objects. However, when it comes to the mitosis and tubule formation scoring and based on the previous observations, an ontological-driven mitosis and tubule formation scoring is highly needed. The results of the image analysis algorithms we

applied emphasize this idea. Like we mentioned previously, the BCGO which we developed describes the histological concepts involved in the breast cancer grading and it drives the semantic annotation and exploration of the image.

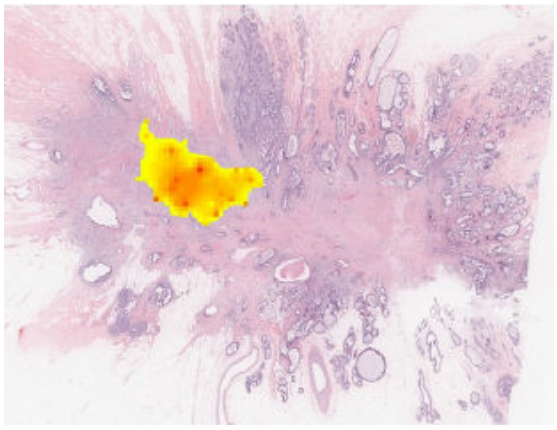


Figure 8.5. WSI global grading using multi-scale dynamic sampling [Veillard et al., 2010]

The mitosis-like objects segmentation is performed using the formal description of the mitosis concept from our ontology through the process of semantic reasoning. Similarly, the tubule structures are identified and annotated based on the knowledge encoded in the ontological description. However, we describe the nuclear pleomorphism criteria in the ontology as well, as to make our ontology complete and useful in other context where it is not directly purposed from image annotation or connection with image analysis level. Figure 8.6 depicts our approach on the mitosis and tubule formation scoring.

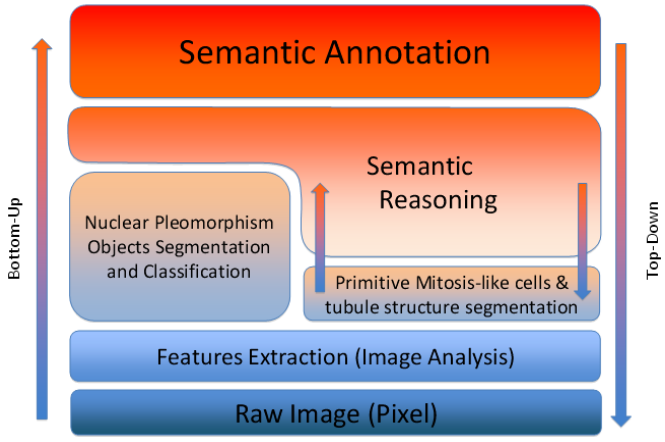


Figure 8.6. Ontology-driven mitosis and tubule formation scoring

The semantic reasoning which is a fundamental part of our formal representation ensures a consistent ontology thus a reliable annotation and retrieval of image in the validation phase. Note that in the image processing and analysis algorithms are given guiding information with regard to ontological concepts. That does not mean they are limited in their own actions. The information captured within ontology is further taken over by the algorithms and used accordingly.
This step of connecting the semantic level with the image level is a very challenging but necessary one in order to fill the semantic gap and to enhance the communication capabilities of the application between the end-user and the machine. Furthermore, this approach comes to narrow the context gap we talked about in chapter 4. By working with semantics in driving the image analysis phase, we are able to obtain a reliable automated breast cancer grading system.

Ontology-based retrieval and validation support

Apart from the novel approaches in image processing and analysis, the virtual cognitive microscope prototype system brings another significant asset: the seamless mapping between the symbolical concepts given by the pathologists and NCI/NIH thesaurus and the spatial representation given by the image analysis algorithms [Roux et al., 2009b].
A typical session with the cognitive virtual microscope makes it possible for the physician to interact with the WSI at different resolutions via the histological concepts represented in the ontology in a graphical-view. If the physician selects a specific concept in the ontology -or by the more sophisticated means of a semantic query (described in chapter 7) - not only the virtual slide shows up the spatial representation of this concept, in terms of instances that belong to that concept, over the WSI but also the real microscope lens can move to the specific location (visual positioning [Begelman, 2006]) in the real histological sample, giving the system capacities to act on the environment. For instance, when we have the query "Find all nuclei that are mitosis", nucleus and mitosis are concept, but the system will retrieve all instances of nucleus that correspond to the criteria of a mitosis. Hence, the pathologist is able to check the consistency of the semantic annotation and to visualize the corresponding frames on the real microscope, before validating the final grading.
Figure 8.7 depicts the annotation of the histological slide based on the concepts from the ontology and Figure 8.8 illustrates the knowledge management part.

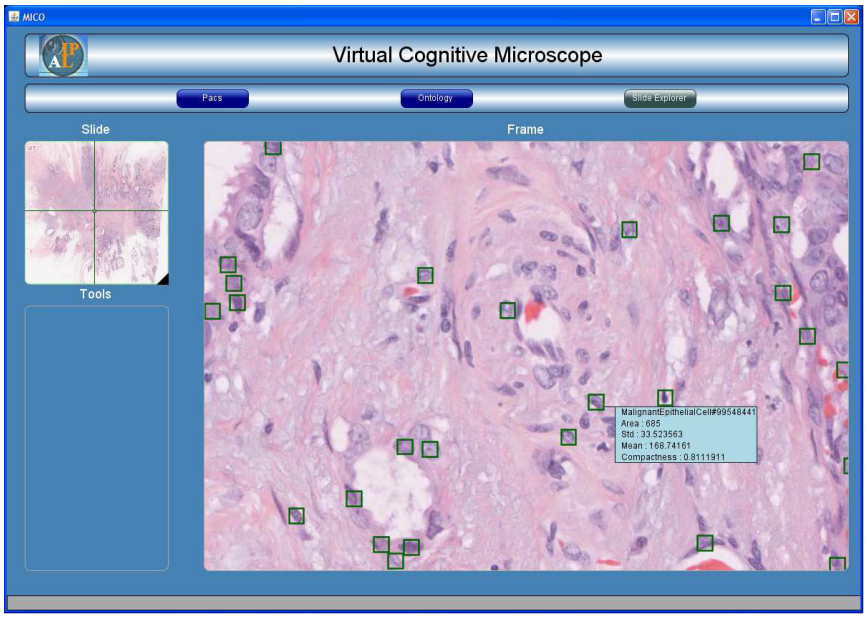


Figure 8.7. Ontology-based annotation and retrieval support

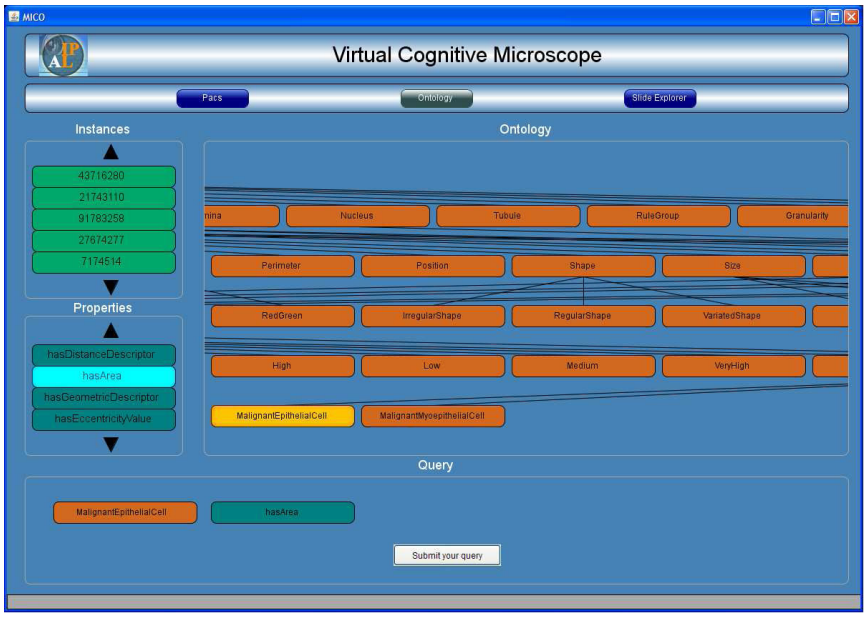


Figure 8.8. Ontology management

CBIR-CBR methodology

In chapter 3 we discussed about an innovative CBIR-CBR hybrid methodology in which ontology and implicitly reasoning, stands as the core of indexing, retrieval and refining steps. In chapter 4 in the image-driven approach, the indexing axis was part of the CBIR-CBR methodology. In this setting, we evaluated six breast core-biopsy cases stained with H&E marker, consisting of 7000 frames scanned from the tumor tissue slides and obtained from the Pathology Department of National University Hospital of Singapore (NUH). The database was composed by two sets: 1400 frames used for the training algorithm phase and 5600 frames used for the testing and validation phase. The slides were scanned on a sequence of frames at 10X40 (400X) magnification with a 1080 X 1024 resolution. Based on the steps discussed in the image-driven approach, the grading was given for each frame and for the entire slide. Individually, the most accurate results were obtained for the mitosis count. Although, a 7, 33 % error was registered for the training dataset and 11% for the testing dataset in a local grading, for the global grading we obtained no computation errors. Compared with the manual grading given by the pathologists, we achieved an accuracy of 80% for the breast cancer global grading. These results were presented in [Tutac et al., 2009a].

Manual Grading				Semi-Automated Grading					
Tubule score	Nuclear score	Mitosis count	BCG	Tubule score	Nuclear score	Mitosis count	BCG	Case ID	Data type
1	1	3	*1*	1	1	3	*1*	1000	Training Database (1400 images)
1	2	1	*1*	2	2	1	*1*	2000	
3	3	3	*3*	3	2	3	*3*	4895	
2	3	3	*3*	3	2	3	*3*	5020	Testing Database (5600 images)
3	3	3	*3*	3	2	3	*3*	5042	
3	2	1	*2*	3	2	1	*2*	5075	

Table 8.2. BCG Grading Results

Data base	Tubule score	Nuclear score	Mitosis count	Component scores error	Global BCG error
Training errors	11%	11%	0	7,33%	0%
Testing errors	11%	22%	0	11%	0%

Table 8.3. Local and global errors

Although we obtained these results, this approach had its drawbacks discussed in chapter 4. Therefore, we took the semantic driven approach as the new direction and what we want to propose with regard to CBIR-CBR is that the MICO system can work within this methodological strategy.

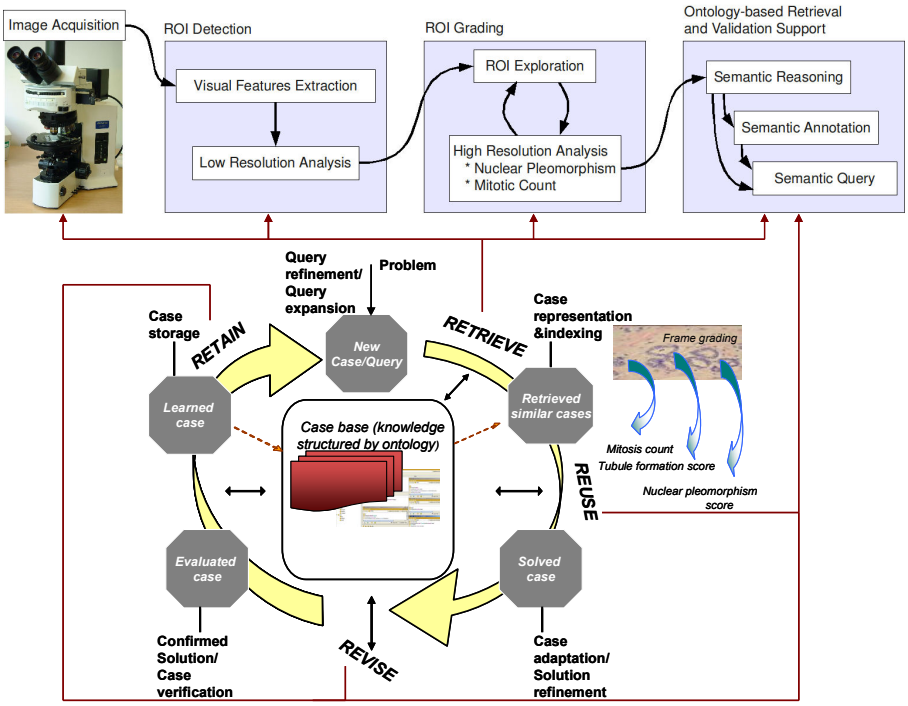


Figure 8.9 MICO prototype based on the CBIR-CBR strategy

Figure 8.9 illustrates our CBIR-CBR envisage on the MICO. The image acquisition, for instance will be performed under a case-type representation and indexing of semantic concepts connected with image features. The ROI detection and multi-scale global nuclear pleomorphism score are the modules which directly work with the image features. The ontology-based retrieval module will follow the *retrieve principle*, which will enable us to find also similar cases to the one submitted as a query. The retrieval results are then used to solve the new case (that is to apply the *reuse principle*). The new knowledge discovery characteristic of the ontology-driven prognosis activates the *revise principle* when the solution for the current case requires some adaptation. Finally, the new knowledge is integrated and the case is stored into the case-base (*the retain principle*). Hence, MICO working by following the principles of CBIR-CBR methodology stands as one of the interesting perspectives for our research team.

8.4. Conclusions

The field of microscopy is experiencing a new shift in paradigms and the need for intelligent interaction in the image exploration tasks is taking very much interest in our days. By fulfilling some of the cognitive aspects of a cognitive vision system as

stated in the roadmap drawn by the European experts, we open interesting avenues for a new microscopic device model coined as the cognitive microscope project. We focused on the histopathological assessment in breast cancer in order to overcome the limitation it currently has. The formal semantic representation encapsulated in the BCGO ontology plays a key role in the MICO prototype as it gives semantic annotation and retrieval support in the grading. We envision that the CBIR-CBR combined strategy could be applied to the MICO system (as shown by Figure 8.9) which is another innovative approach. However, this strategy must be effectively applied and evaluated and this stands as a matter of future work. Together with the other mentioned aspects, this carries an important note even more now that MICO has been recently granted by the Technologies for Health program of the French National Research Agency (ANR).

We believe that the universe of microscopic images can be a very useful research playground to experiment new challenges and ideas in the field of cognitive enhanced vision systems, the microscope being considered as a robot acting on its environment under the supervision of the end-user. A major research issue we aim at is related to continuous learning based on the monitoring of typical session of a diagnosis exploration by histopathological experts.

Hence, the contributions of this chapter are:

- integration of BCGO in the cognitive microscope framework prototype as providing support for semantic annotation and retrieval of histopathologic images
- MICO as virtual breast cancer grading consultant and assistant
- MICO as seen from the cognitive perspective
- CBIR-CBR hybrid methodology applied on the MICO system

9. Conclusions and Perspectives

9.1. Contributions Summary

The subject of this thesis is the formal representation and reasoning in medical applications. Our personal take on this matter is the attempt to overcome several problems in microscopic image-based prognosis exemplified by the application of breast cancer grading. In point of fact, the underlying idea is to bring together the scientific community and the medical field by means of ontological representation and formal logic reasoning in the breast cancer grading domain knowledge.

The state-of-the-art literature on representation and reasoning concerning images and concepts introduced in the first three chapters of the thesis revealed that:

- a bridge between image representation and formal concept representation is highly needed in medical applications. This bridge addresses the semantic gap and context gap problems commonly encountered in this field.
- qualitative representations are more appropriate in the context of medical applications since diagnosis and prognosis are subject to personal experience and stamina of each individual pathologist although there are standardized guidelines which provide quantitative facts to direct the assessments.
- content-based image approaches are necessary but not sufficient for indexing, retrieval and especially refining processes. In the same vein, approaches that resemble with medical procedure and reasoning but without semantics are not fully reliable.
- high expressivity coupled with decidability power is strongly demanded to provide efficient formal representation and reasoning. To this end ontological approach with DL reasoning meet the purpose. Additionally, rules bring higher expressivity yet to the expense of decidability. Therefore, if rules are needed, DL safe rules must be constructed.
- since images deal with spatial objects and their relations, a spatial representation is essential to complete the formal knowledge representation.

Based on these results of the studies, the main contributions we bring in this thesis are the following:

- *a comparative analysis from a methodology and not from technology perspective of two paradigms from different fields, CBIR and CBR taken further to medical applications (chapter 2 and 3).*

We present the definition and the specific features of each paradigm, and we argue that in order to do a reliable comparative analysis of them, they must be considered at the same level and that is the level of methodology [Tutac et al., 2009a]. It follows that indexing, retrieval and refining which are techniques encountered in both, can now be examined in detail. We then extend the comparison to their usage in clinical applications. Our purpose is to set a foundation based on a hybrid CBIR-CBR approach by taking the advantages of both into a single framework (with CBIR as a subset of CBR). We envision this strategy can be used in the cognitive virtual microscope system as support for indexing, retrieval and refining process.

- *an approach to bring together image representation and concept representation as both employ knowledge representation and both concern perception and cognition (chapter 4)*.

This consists of a novel theoretical formal model for the knowlege representation and reasoning in the domain of breast cancer grading: the breast cancer grading application ontology BCGO based on a qualitative modeling philosophy of perdurants [Roux et al., 2009a], [Tutac et al., 2009d].
A discussion over qualitative versus quantitative representation is fundamentally important for the building of the formal model. The type of knowledge captured in the formal model is also established as being the knowledge which describes perdurants. The rationale is that the characteristic of breast cancer grading is the analysis of what can be seen under the microscope at the current moment.
Based on the demarche from CBIR and CBR to ontologies (chapter 3), we propose two different novel approaches: the image-driven and the concept-driven approaches in which we show our solutions to narrow the semantic and the context gaps discussed in chapter 2 and 3 [Tutac et al., 2008a], [Tutac et al., 2008b], [Dalle et al., 2008]. The results of their analysis show that the concept-driven is more appropiate for the objective of a formal representation and reasoning in breast cancer grading domain. It also shows that the reasoning is the cue of each paradigm. The assets of this model are manifold:

- bridges the *semantic gap and context gap* by a formal logic *semantic indexing* technique.

Ontologies are by now commonly accepted to be a solution for the semantic gap. In our approach, this fact is taken further from the theoretical level to the application level. The interconnection between image and semantics is illustrated in the cognitive microscope framework developed by our team. We propose a method to link the formal language with computer programming languages in order to convey the semantics and the formal description to the image analysis algorithms (chapter 8). The fact that this model can be applied as well in other histological grading is an important contribuition to solving the context problem.

- offers h*igh expressivity and decidability* powers provided by structural modeling (OWL-DL) and rule modeling (SWRL) modules of the BCGO [Roux et al., 2009a], [Tutac et al., 2009d].

Apart from the fact that an ontological representation of the breast cancer grading domain has not been proposed elsewhere in the literature, the combined approach

of OWL-DL with SWRL is another novelty in this domain. The tradeoff between expressivity and decidability is discussed in chapter 6 where we show how to create SWRL DL safe rules in our application ontology, when the need for DL safe rules is at hand. The implementation of the ontology (chapter 6) is not merely a cronological description of classes, or properties but it revolves around the most important characterististics of OWL-DL and SWRL languages and treats the pitfalls of each of them.
A standard ontological representation is a very helpful tool when performing the prognosis assessment. It complements the standard grading system (NGS) with terminological consistent description and furthermore as it drives the image exploration phase, it identifies the specific objects such as the tubular structures, mitotic figures which the pathologists manually look for in a standard procedure. It thus serves as an automated second opinion or a virtual consultant for the grading.

- *semantic reasoning* with *spatial reasoning* based on a formal spatial theory and *DL tableau-algorithm* [Tutac et al., 2009b], [Tutac et al., 2009c] , [Tutac et al., 2010a-b]. (*chapter 5*).

The semantic reasoning is a qualitative reasoning based on the fact that the representation itself is qualitative. The formal spatial theory approach presents four types of spatial concepts: mereo-topologic, metric, geometric and dimensional concepts. The focus is however set on the mereo-topologic and metric spatial concepts. We take two different yet connected approaches for the representation of these spatial concepts. The common ground is the region connection calculus and the composition table. How this approach works is clearly illustrated for the *CloseTo* metric relation.
Using a formal theory support has the advantage of eliminating ambiguities or inconsistencies in the representation (as illustrated for inclusion relation). Another advantage consists of defining more complex relations such as a relation that depicts a region located in another region but which is not a proper part of that region.
The qualitative reasoning we propose is twofold: manual and automated. The manual reasoning follows the reasoning of the humans- the medical experts and we show how the theorems and properties of mereo-topology and metric apply to BCG. The automated reasoning consists of describing the tableau-based algorithm which is used in the implementation of the Pellet reasoner. The concept satisfiability tasks to which all other reasoning tasks can be reduced (as discussed in the literaturre) is illustrated by using an example from our BCG ontology. In DL reasoning terms, a model is obtained if the algorithm found consistency in the interpretation of the representation.

- *semantic retrieval for evaluation and validation of the BCGO (chapter 7).*

Our BCGO has 129 classes, based on the hierarchy of four main classes, AnatomicalEntity, ConceptualEntity, SpatialEntity and MicroscopicEntity. A number of 169 instances of these classes were generated based on is-a relationship and 86 properties were defined to capture the relationships between instances and between classes. In terms of levels of hierarchy we have a depth of 10, with an average of 3 siblings and maximum 4 siblings. The total number of restriction is 138, with 79

existential, 49 universal, 1 cardinality restriction, 2 max cardinality restriction and 1 hasValue property.
The evaluation is performed from the granularity, degree of reflection and integration perspectives. These metrics are relevant in the context of a qualitative representation and moreover of an application ontology. Another important aspect related to the evaluation concerns is the syntactic constraint of OWL-DL with respect to OWL-Full. In the validation phase, the contribution we bring is a combined semantic retrieval with medical and ontological community feedback. We show how semantic queries of different types (e.g. SQWRL, RDF) help the computer scientist as well as the pathologist to extract information from the ontology and to further validate the representation. There are several reasons to why semantic queries are useful: the medical community is using formalized structured representation for the medical knowledge, a computer-aided diagnosis system with this kind of implementation can perform the grading to give a second opinion and overcome the time problem of a manual grading and lastly, they are useful to the efficiency of the consensus meetings of the pathologists, as they have them not very often (due to time and amount of work constraints).

- *BCGO integration in a virtual microscope framework (chapter 8).*

This aspect gives a significant cognitive value to the framework as emphasized by the MICO prototype system [Roux et al., 2009a], [Tutac et al., 2009c], [Roux et al., 2009b], [Loménie et al., 2009] built in our previous work [Racoceanu et al., 2009]. The most important contribution is represented by the ontology-driven mitosis and tubule formation approach as well as the ontological validation support. Furthermore, the CBIR-CBR hybrid framework is another asset for the indexing, retrieval and refining processes in the MICO system [Tutac et al., 2009a].

9.2. Research Perspectives

There are several directions that could be further investigated. We firstly consider the refining and the evaluation of the ontology.
The refining of the ontology implies adding new knowledge such as scale information, since the scale plays an important role in the histopathological grading. The criteria of breast cancer grading specify the magnification for acquisition of tubule structures, or mitosis figures. To this end, the spatial theory for the dimensional relation may be further extended. Additionally, the spatial theory could be extended with spatial representation of the geometric relations. Although our approach is a qualitative one, an aspect which we emphasize along the thesis, of significant note is the need for a quantitative evaluation of such theory of spatial relations, such that its advantage would clearly transpire.
Another fact that refining of the ontology implies is to produce even more accurate representation of the knowledge from the medical perspective. The medical feedback is very important in this process. As we mentioned in chapter 7, further evaluations on the applicability of our approach in a clinical setting in terms of quality enhancement and time reducing in the work of the pathologists, are however

needed. Ultimately, refining deals with conforming the representation to the DL characteristics such that the ontology remains consistent.

Evaluation metrics may differ from the type of the ontology to another. Therefore, sometimes it is not relevant to apply the same metrics proposed for reference ontology to application ontology. However, it is important to firstly find generic metrics (as we discussed in chapter 7) and apply them to the new ontology built. If there is a purpose of later integration of an application ontology into a reference ontology in mind, then the metrics proposed for reference ontology may be useful. As specified in chapter 7, another objective is to integrate the BCGO into reference ontology (which can be found in OBO) that handles concepts from the breast cancer domain. This is the reason we mention that OBO also uses some simple metrics to evaluate the ontologies from their framework.
The segmentation of ontologies and evaluation of each segment provides more efficient and optimized results than evaluating large-sized reference ontology as presented by [Seidenberg and Rector, 2006]. We also said in chapter 4 that our method of segmentation could be further improved by applying some of the methods proposed in the above mentioned paper.
Another way of evaluation is by comparison of same type ontological representation. In our particular case, there is no other application ontology that describes the breast cancer grading. Furthermore, application ontologies convey the knowledge from a specific domain, but the domain can be larger from one context to another. For instance, the breast cancer grading information which is represented in the BCGO may be less than the FMA-Radlex ontology due to the dimension of the corpus of information and also to how much of other related-knowledge the curators want to integrate. The BCGO can also integrate information related to other prognostic or diagnostic factors such as: staging, hormones and HER receptors. The question is then, how relevant the metric of class richness is (or any other metric), as the number of classes may differ from one application ontology to another. Consequently, a direct comparison of two different ontologies although they both may be of the same type, does not provide relevant results. To this end, applying other evaluation metrics still remains a challenge for future work.
In the context of the CBIR-CBR methodology for the MICO framework, another direction of evaluation would be needed: a comparative analysis between the manual grading of the pathologists, automated grading based only on image processing algorithms and grading based on the ontology guidance. We envision the procedure for this work would follow three steps:

- defining a methodology for an evaluation of this type
- defining a set of metrics
- applying the metrics

At the current stage of our work, this would have been very difficult for multiple reasons. Firstly, a manual annotation of all specific objects from the histological image and grading of all the frames and slides is needed in order to have a ground truth to further compare with. However, this is time consuming and a tedious work for pathologists. On a daily basis, they evaluate around 200 cases/slides in Singapore University Hospital and 900 cases/slides in Pitié-Salpêtrière Hospital from Paris (the partner in our ANR project). This implies around 4000 frames/slide. In our database we have 20 slides which give us 80.000 frames that require manual annotation and grading. Therefore, we considered that producing an automated

ground truth is highly needed. In this line, one of the purposes of MICO is to be a ground truth maker for histopathology breast cancer biopsies in which ontology plays an importance role for the annotation and exploration of the slides. However, in this context, a classical CBIR method in which a histological image is compared against the database images to give evaluation results with precision and recall metrics is not useful (as discussed in chapter 7). And this represents the third reason for the difficulty of such an evaluation. Nevertheless, this constitutes a very important direction of the MICO project.

The issue of mapping and alignment is strongly related to the integration of application ontologies into reference ontologies. However, this is a complex topic by itself. A state-of-the-art and a theoretical approach for defining ontology mapping are presented in [Kalfoglou and Schorlemmer, 2003].

Another direction concerns the complexity of the reasoning algorithm, especially if there are some non-deterministic representation (e.g. when we have to apply union rules to construct the search path of a model). In the worst case, a OWL-DL representation which corresponds to SHOIN (D) language is of EXPtime complexity. The issue of large *AB*oxes is also important in the context of algorithm complexity. Our ontology is an application ontology hence this aspect did not hold interest in our current approach. There are also several other aspects related to optimization or the tracing of the tableau-algorithm (generating explanations for the inference process, inconsistency description to the user) which can be further discussed in the context of our ontology as approached by [Kalyanpur et al, 2005]. Testing our ontology in SWOOP[12], the hypermedia OWL ontology editor, can also provide useful analysis and further improvement. Additionally, various tests and experiments using different tableau-based reasoners such as FACT++ or RacerPro, can be carried our on the BCG ontology. However, the study conducted by [Sirin et al., 2005] emphasizes that Pellet reasoner has unique functionalities, such as reasoning with individuals (nominals and conjunctive *AB*ox queries) and reveals that its performance ranges from acceptable to very good performance for varying complexity and expressivity of ontologies.

Encapsulating histopathology data in an ontological model could be applied to other diseases such as prostate cancer. Alternatively, it can help in discovering new knowledge by means of ontological -driven image exploration as this is one of our immediate goals in the cognitive microscope framework. In this light, the spatial theory support introduced in chapter 5 gives another future direction for research. It can be refined and extended as well, to incorporate other spatial relations which will help in the discovering of new knowledge. Another idea is to propose other approaches for spatial theory. The hybrid method of RCC-8 and OWL/RDF reasoning discussed in [Stocker and Sirin, 2009] advances a spatial Pellet DL reasoner which opens promising perspectives. In the context of the cognitive microscope framework, we mentioned in chapter 8, that we envision CBIR-CBR combined strategy could be applied to the MICO system. However, as we also said, this strategy must be effectively applied and evaluated and this stands as a matter of future work.

[12] http://www.mindswap.org/2004/SWOOP/

Our ontology can be integrated into a complex framework that handles multiple models and data such as system biology models from molecular and cellular level and data from cytopathology or radiology level. Related works are found in the Virtual Physiological Human (VPH) community, where biomedical modeling and in-silico (computerized) simulation of human body challenge the improvement of the ability to predict, diagnose and treat disease, and also the impact on the future of healthcare, the pharmaceutical and medical device industries[13]. The extension of representation from tissue to cellular or molecular level or even to the organ level integrated into a single formal representation comes to meet the need of medical experts to benefit of low-level and high-level information at the same time.
An illustrative example is the recently launched RICORDO[14] project financed by the FP7 European programme. The project uses ontology-based metadata composite annotation to describe a large corpus of data types and models in order to achieve interoperability. Another project funded by FP7 programme called Debug-IT[15] [Lovis et al., 2009] challenges the same idea of improving the healthcare by sharing heterogeneous clinical data sets from different hospitals in different countries, with different languages and legislations. It also relies on ontologies as it is described in [Ouagne, 2009].

Ontology-driven diagnosis or prognosis is another direction we envision to be of high relevance in the future, as ontologies are the heart of the semantic web or the nucleus of web cells if we speak from the histopathological perspective. Another significant aspect here is that ontology can contribute to reducing the inter-observer agreement inconsistencies as proven by [Steichen et al., 2006]. This is a matter of interest for us as in the MICO project granted by ANR, IPAL collaborates with Pitié-Salpêtrière Hospital from Paris. New systems for prognosis, the telepathology for instance, or computer-aided diagnosis systems can be built using the philosophy of ontology-driven approach. The cognitive microscope framework initiative proposed by our IPAL lab emphasizes this very idea. The VPH community provides such opportunities since its vision is to develop ICT-based tools for modeling and simulation of human physiology and disease-related processes.

In this train of challenging perspectives when speaking of ontology reusability, the integration in larger reference ontology or representation of heterogeneous data from different fields, it then becomes essential if the desideratum of decidability of reasoning algorithm still holds. And since integration of ontology and an extended representation with concepts from other fields, is also of future work for us, the issue of OWL-DL paradigm and OWL-Full carries significant value. In a similar setting (with many various data), the philosophy of reasoning in AGFA Healthcare strategies relies on a constraint-free first-order logics approach such as OWL-Full to capture expressivity at the highest level [Lovis et al., 2009]. The time optimization when computing the classification and checking the consistency of the ontology is then solved using a backward-chaining reasoner enhanced with Euler mechanism of path detection developed by AGFA.

Lastly, another direction could be a temporal extension of the formal representation and reasoning, besides the spatial. In this way, we could model perdurants

[13] http://www.vph-noe.eu/
[14] http://www.ricordo.eu/
[15] http://www.debugit.eu/about/introduction.html

(processes) in our ontology in a dynamic fashion (e.g. capturing the process of a nucleus becoming a mitosis, or follow-ups of the breast cancer grading). This is what we called morphogenesis in chapter 4. Our interest stems from the molecular/cellular formal representation using SBML mark-up language[16] to capture time dimension which is also advanced by VPH community. Additionally, the temporal extension of DL conveys the same idea of temporal information integration into an ontological representation.

[16] http://sbml.org/About, last accessed July 2010

Bibliography

[Aamodt and Plaza, 1994] A. Aamodt and R. Plaza "Case Based Reasoning: Foundational Issues, Methodological Variations and System Approaches", *AI Communications*, vol.7, no.1, pp. 39-59, 1994.

[Adawi et al., 2006] M. Adawi, Z. Shehab, H. Keshk and M. Shourbagy, "A fast algorithm for segmentation of microscopic cell images", in *Proc. 4th Int. Conf. Inf. & Com. Tec*, 2006.

[Aiello, 2002] M.Aiello, "Spatial reasoning Theory and Practice", PhD thesis, 2002

[Artale and Franconi, 2001] A. Artale and E.Franconi, "A survey of temporal extensions of description logics", *Annals of Mathematics and Artificial Intelligence*, Kluwer Academic publishers, vol.30, pp.171-210, 2001.

[Artale et al., 2007] A. Artale, D.Calvanese, R.Kontchakov and M.Zakharyaschev, "DL-Lite in the light on first-order logic", in *Proc of 22nd national conference on Artificial Intelligence*, vol.1, pp. 361-366, 2007.

[Artale et al., 2008] A. Artale, E.Franconi, F.Wolter and M.Zakharyaschev, "A temporal description logics for reasoning over conceptual schemas and queries", *Lecture Notes in Computer Science*, vol. 2424, 2002.

[Auer et al., 2005] ECVision: The European Research Network for Cognitive Computer Vision Systems, "A Research Roadmap of Cognitive Vision", IST Project IST-2001-35454, 2005.

[Baader et al, 2003a] F. Baader, D.L. McGuiness, D. Nardi, and P.F. Patel-Schneider, *The Description Logic Handbook: theory, implementation and, applications*, Cambridge Univ Press, 2003.

[Baader et al., 2003b] F. Baader, I.Horrocks, and U. Sattler, "Description Logics as Ontology Languages for the Semantic Web", in *Lecture Notes in Artificial Intelligence*, Springer-Verlag Ed, pp. 228-248, 2003.

[Baader et al., 2007] F. Baader, D.L. McGuiness, D. Nardi, and P.F. Patel-Schneider, *The Description Logic Handbook: theory, implementation and, applications*, Cambridge Univ Press, 2nd ed., 2007.

[Baader and Sattler, 2000] F.Baader and U. Sattler, "An overview of tableau algorithms for Description Logics", *Studia Logica,* vol.69, no.1, pp. 5-40, 2000.

[Bateman and Farrar, 2005] J. Bateman and S. Farrar, "Spatial Ontoloy Baseline", Tech. Rep, Bremen, SFB/TR8, 2005.

[Baumeister and Seipel, 2005] J. Baumeister and D. Seipel, "Smelly OWLs- Design Anomalies in ontologies", *Proc. of the 18th International Florida Artificial Intelligence Research Society Conference, AAAI Press*, pp. 284, 2005.

[Beliën et al., 1997] A. Beliën, J. Baak, P. van Diest and A. Ginkel, "Counting mitosis by image processing in feulgen stained breast cancer sections: the influence of resolution", *Cytometry,* vol.28 no.2, pp.135- 140, 1997.

[Begelman, 2006] G. Begelman, M. Lifshits and E. Rivlin, "Visual Positioning of Previously Defined ROIs on Microscopic Slides*", IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no.1, 2006.

[Berner, 1999] E.S. Berner, *Clinical Decision Support Systems: theory and practice*, Springer-Verlag, New York, pp.77-99, 1999.

[Bichindaritz, 2003] I. Bichindaritz, "Solving safety implications in case-based decision support system in medicine", *Workshop on CBR in the Health Sciences*, pp. 9-18, 2003.

[Bichindaritz and Marling, 2006] I.Bichindaritz and C.Marling, "Case-based reasoning in the health sciences: What's next?", *Artificial Intelligence in Medicine*, vol.36, no.2, pp.127-135, 2006.

[Bodenreider and Burgun, 2005] O. Bodenreider and A.Burgun, "Biomedical Ontologies", in *Medical informatics: knowledge management and data mining in biomedicine*, Chen, H., Fuller, S.S., Friedman, C., Hersh, W. (eds), ISBN: 0-387-24381-X, Springer, pp. 211-234, 2005.

[Bodenreider and Zhang, 2006] O. Bodenreider, and S. Zhang, "Comparing the representation of anatomy in the FMA and SNOMED-CT", *AMIA Annual Symposium Proc.,* pp. 46-50, 2006.

[Boley et al., 2005] H.Boley, B.Grosof and S.Tabet, "RuleML tutorial", available at http://www.ruleml.org/papers/tutorial-ruleml-20050513.html, 2005, last accessed July 2010.

[Bontas et al., 2004] E.P. Bontas, R. Tolksdorf and T. Schrader, "Ontology-based Knowledge Organization in a Semantic Web for Pathology," in Proc. Inquiring Knowledge Networks on the Web Conference, IKNOW 2004.

[Brageul and Guesgen, 2007] D. Brageul and H. Guesgen, "A model for qualitative spatial reasoning combining topology, orientation and distance", *Proc. AAAI*, pp. 653-658, 2007.

[Brank et al., 2005] J.Brank, M. Grobenlink and D.Mladenic, "A Survey of Ontology Evaluation Techniques", in *Proc. Data Mining and Data Warehouses* SiKDD, 2005.

[Burrieza et al., 2009] A. Burrieza, E. Munoz-Velasco and M.Ojeda-Aciego, "Closeness and Distance relations in order of magnitude qualitative reasoning via PDL", *Qualitative Reasoning Workshop*, 2009.

[Carneiro et al., 2007] G. Carneiro, A. Chan., P. Moreno and N. Vasconcelos, "Supervised Learning on Semantic Classes for Image Annotation and Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 vol.3, pp.394-410, 2007.

[Cardiff and Jensen, 2000] R.Cardiff and R. Jensen, Histological Grading of Breast Cancer, http://ccm.ucdavis.edu/bcancercd/311/grading_diagram.html, 2000, last accessed July 2010.

[Chandrasekaran et al., 1999] B. Chandrasekaran, J. Josephson, R. Benjamins, "What are Ontologies and why do we need them*?",* IEEE Intelligent Systems,http://www.cse.ohio-state.edu/~chandra/What-are-ontologies-and-why-we-need-them.pdf, 1999, last accessed July 2010.

[Checkland and Scholes, 1990] P. Checkland and J. Scholes, *Soft Systems Methodology in Action*, Wiley, NY, 1990.

[Cohn and Renz, 2008] A. Cohn and J. Renz, "Qualitative spatial representation and reasoning", in: F. van Hermelen, V. Lifschitz, B. Porter, eds., *Handbook of Knowledge Representation*, Elsevier, 551-596, 2008.

[Corcho et al., 2004] O.Corcho, A.Gomez-Perez, R.Gonzalez-Cabero and C. Suarez-Figueroa, "ODEval: A tool for evaluating RDF(s), DAML + OIL, And OWL Concept Taxonomies", pp 1-13, 2004.

[Dalle et al., 2008] J.Dalle, D. Racoceanu, W.K Leow**,** A.E. Tutac and T.Putti, "Automatic Breast Cancer Grading of Histopathological Images", *Proc of 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* IEEE EMBS, ISBN 978-1-4244-1814-5, ISSN 1557-170X, pp.3052-3055, Vancouver, Canada, 2008.

[Dalle et al., 2009] J.R. Dalle, H. Li, C.-H. Huang, W.K. Leow, D. Racoceanu and T. C. Putti, "Nuclear pleomorphism scoring by selective cell nuclei detection," *IEEE Workshop on Applications of Computer Vision* (WACV 2009), Snowbird, Utah, USA, 2009.

[Dalton et al., 2000] L.W. Dalton, S. Pinder, C. Elston, I. Ellis, D. Page, W. Dupont and R. Blamey, "Histological Grading of Breast Cancer: Linkage of Patient Outcome with level of pathologist agreement", Journal of Modern Pathology, vol.13, no.7, pp.730-735, 2000.

[Damjanović et al., 2003] V. Damjanović, D. Gašević, and V. Devedžić, "Ontology Validation", in *Proc. 6th International Conference of Information Technology*, Bhubaneswar, pp.183-186, 2003.

[Dasmahapatra et O'Hara, 2006] S. Dasmahapatra and K.O'Hara, "Interpretation of Ontologies for Breast Cancer", Triple C, vol. 4, no.2, pp.293-303, 2006.

[Datta et al., 2008] R. Datta, D. Joshi, J. Li and J. Wang, "Image Retrieval: Ideas, Influences, and Trends of New Age", *ACM Transactions on Computing Surveys*, vol.40, no.2, pp.1-66, 2008.

[Davis et al., 1993] R. Davis, H. Shrobe and P. Szolovits, "What is a Knowledge Representation?" *AI Magazine,* vol. 14, no.1, pp.17-33, 1993.

[Demir and Yener, 2005] C. Demir and B. Yener, Automatic cancer diagnosis based on histopathological images: a sistematic survey, Technical Report, 1-16, TR 05-09, 2005.

[Deselaers and Müller, 2005] T. Deselaers and H. Müller, Tutorial in relevance feedback, http://thomas.deselaers.de.relevance-feedback.pdf, 2005, accessed in 2008.

[Deserno et al., 2007] T. Deserno, S. Antani and R.Long, "Gaps in content-based image retrieval", *Proc SPIE*, vol. 6516, pp. 1-11, 2007.

[Donnelli et al., 2005] M. Donnelli, T. Bittner, and C. Rosse, "A formal theory for spatial representation and reasoning in biomedical ontologies", *Artificial Intelligence in Medicine,* vol. 36, pp. 1-27, Jul. 2005.

[Drummond and Shearer, 2006] N.Drummond and R. Shearer, "Open World Assumption", OWA.pdf, 2006, last accessed 2010.

[Fernandez et al., 2006] M.Fernandez, I.Cantador and P.Castells, "CORE: A Tool for Collaborative Ontology Reuse and Evaluation", *Proc. 4th Int. Workshop on Evaluation of Ontologies for the Web* (EON'06), at the 15th Int. WorldWide Web Conference (WWW'06). Edinburgh, UK, 2006.

[Foy et al., 2007] N.Foy, S. de Coronado, H.Solbrig, G. Fragoso, F.Hartel and M.Musen, "Representing the NCI thesaurus in OWL-DL: modeling tools help modeling languages", *Applied Ontology*, pp.1-17, IOS press, 2007.

[Freksa, 1991] C. Freksa, "Qualitative Spatial Reasoning", *Cognitive and Linguistic Aspects of Geographic Space*, D.M. Mark & A.U. Frank eds., pp. 361-372, 1991.

[Frkovic-Grazio and Bracko, 2002] S. Frkovic-Grazio, and M. Bracko, "Long term prognostic value of Nottingham histological grade and its components in early (pT1N0M0) breast carcinoma," *Journal of Clinical Pathology*, vol. 55, no. 2, pp. 88–92, 2002.

[Gangemi et al., 2005] A.Gangemi, C.Catenacci, M.Ciaramita and J.Lehmann, "A theoretical framework for ontology evaluation and validation", *Proc. SWAP*, 2005.

[Golbeck et al., 2003] J. Golbeck, G. Fragoso, F. Hartel, J. Hendler, J.Oberthaler, and B. Parsia, "The National Cancer Institute's Thesaurus and Ontology", *Journal of Web Semantics,* pp.75-80, Jul. 2003.

[Golbreich et al., 2005] C. Golbreich, O.Bierlaire, O. Dameron and B. Gibaud, "What reasoning support for Ontology and Rules? the brain anatomy case study", *Proc. 8th International Protégé*, pp. 1-9, 2005.

[Grau et al., 2008] B. Cuenca Grau, I. Horrocks, B. Motik, B, Parsia, P. Patel-Schneider and U. Sattler, "OWL 2 : The next step for OWL", Web Semantics: Science, *Services and Agents on the World Wide Web*, vol.6, no.4, pp.309-322, 2008.

[Grenon and Smith, 2004] P. Grenon, and B. Smith, "SNAP and SPAN: Toward dynamic spatial ontology", *Spatial Cognition and Computation,* vol. 4, no. 1. pp. 69-103, 2004.

[Gruber, 1993] T. Gruber, "A Translation Approach to Portable Ontology Specification", *KA*, vol. 5, no. 2, 199-220, 1993.

[Gruber, 1995] T. Gruber, "Toward Principles for the Design of Ontologies used for Knowledge Sharing," *International Journal of Human-Computer Studies*, vol. 43, no. 5-6, pp. 907–928, Nov. 1995.

[Guizzardi, 2005] G.Guizzardi, "Ontological Foundations for Structural Conceptual Models", Centre for Telematics and Information Technology CTIT PhD Thesis Series, no.05-74, Telematica Instituut Fundamental Research Series, no.15, 2005.

[Hanby, 2005] A.Hanby, "The pathology of breast cancer and the role of the histopathology laboratory", Journal of Clinical Oncology, vol.17, no.4, pp. 234-239, 2005.

[Herchenröder, 2006] T.Herchenröder, "Lightweight semantic web oriented reasoning in Prolog: tableaux inference for description logics", Master of science, School of Informatics, University of Edinburgh, 2006.

[Hidki et al., 2007] A. Hidki, A.Depeursinge, J.Lavindrasana, M. Pitkanen., X. Zhou and H. Müller., "The medGIFT project: perspective of a medical doctor", *Journal of Medical Imaging Technology*, vol. 25, no. 5, pp. 356-361, 2007.

[Holt et al., 2006] A. Holt, I. Bichindaritz, R. Schmidt and P. Perner, "Medical applications in case-based reasoning", *The Knowledge Engineering Review*, vol. 20 no. 3, pp. 289-292, 2006.

[Horrocks, 2000] I.Horrocks, "Description Logic Reasoning", presented at the *International Conf on Logic Programming and Automated Reasoning*, Montevideo, Uruguay, 2000.

[Horrocks et al., 2003] I. Horrocks, P.F. Patel-Schneider, and F. van Harmelen, "From SHIQ and RDF to OWL: The making of a web ontology language", *Journal of Web Semantics,* vol.1, no.1, pp. 7-26, 2003.

[Horrocks et al., 2004] I.Horocks, P.F.Patel-Schneider, H.Boley, S.Tabet, B.Grosof and M.Dean, "SWRL: A Semantic Web Language combining OWL and RuleML", available at http://www.w3.org/Submission/SWRL/, 2004, last accessed July 2010.

[Horrocks et al., 2007] I. Horrocks, P. F. Patel-Schneider, D. L. McGuinness, and C. A. Welty. OWL: A Description Logic Based Ontology Language for the Semantic Web; Deborah L. McGuinness and Peter F. Patel-Schneider, "From Description Logic Provers to Knowledge Representation Systems", pp. 458—486. In *The Description Logic Handbook: Theory, Implementation and Applications*, ed. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider. Cambridge University Press, 2nd edition, August 2007.

[Huang et al., 2010] C-H. Huang, D. Racoceanu, L. Roux and T. Putti, "Bio-inspired Computer Visual System using GPU and Visual Pattern Assessment Language (ViPAL): Application on Breast Cancer Prognosis*", IJCNN,* Barcelona, Spain, July 18-23, 2010.

[Hudelot et al., 2006] C. Hudelot, J. Atif and I. Bloch, "Ontologies de relations spatiales floues pour l'interprétation d'images", in *Rencontres francophones sur la logique floue et ses applications*, Toulouse, pp. 363-370, 2006.

[Hudelot et al., 2008] C.Hudelot, J.Atif and I.Bloch, "A spatial relation ontology using mathematical morphology and description logics for spatial reasoning", ECAI workshop on Spatial and Temporal Reasoning, pp.21-25, July.2008.

[Iskandar et al., 2007] A.Iskandar, J..Thom and S.Tahaghoghi, "Querying image ontology", Proc. Australasian Document Computing Symposium, pp.84-87, 2007.

[Jeong et al., 2005] H. Jeong, T. Kim, H. Hwang and H-J. Choi, "Comparison of thresholding methods for breast tumor cells segmentation", in *Proc of 7th Int. Workshop on Enterprise networking and computing in Healthcare Industry,* pp. 392-395, 2005.

[Jurisica et al., 2001] I.Jurisica, P.Rogers, J.Glasgow, S.Fortier, J.Luft, M.Bianca, R.Collins and G.de Titta, "Image feature extraction for protein crystallization: integrating image analysis into case-based reasoning", *Proc. IAAI*, pp. 1-8, 2001.

[Kadijevic, 2002] D. Kadijevic, "Are quantitative and qualitative reasoning related? A ninth grade pilot study on multiple proportion", *The Teaching of Mathematics*, vol.5, no.2, pp.91-98, 2002.

[Kalfoglou and Schorlemmer, 2003] Y. Kalfoglou and M.Schorlemmer, "Ontology mapping: the state of the art", *The Knowledge Engineering review*, vol.18, no.1, pp.1-31, 2003.

[Kalfoglou et al., 2006] Y. Kalfoglou, S. Dasmahapatra, D.Dupplow, B.Hu, P.Lewis, and N.Shadbolt, "Living with the semantic gap: Experiences and remedies in the context of medical imaging", *1st International Conference on Semantics and Digital Media Technologies*, pp.46-47, 2006.

[Kalyanpur et al, 2005] A. Kalyanpur, B.Parsia, B.C. Grau and E.Sirin, "Tableau tracing in SHOIN", Technical report UMIACS-TR 2005-66, pp.1-28, 2005.

[Kamp et al., 1998] G. Kamp, S. Lange and C. Globig, *Case-based Reasoning Technology: from Foundations to Application*, M. Lenz, Springer, Berlin, 327, 1998.

[Karimi, 2008] V.Karimi, "Semantic Web Rule Language", http://www.cs.uwaterloo.ca/~gweddell/cs848/Vahid.pdf, 2008, last accessed July 2010.

[Knublauch et al., 2004] H.Knublauch, R.Fergerson, N.Foy and M.Musen, "The Protégé-OWL plugin: an ppen development environment for semantic web applications", 3rd International Semantic Web Conference, Japan, pp. 1-14, 2004.

[Kolodner, 1993] J.Kolodner, *Case-Based Reasoning*, Morgan Kaufmann, San Francisco, 1993.

[Koubarakis, 2010] M. Koubarakis, "Table proof techniques for DLs", Knowledge Technologies, tableaux-techniques1spp.pdf, accessed 2010.

[Lamard et al., 2007] M.Lamard, G.Cazuguel, Q.Quellec, L.Bekri C.Roux and B.Cochener, "Content-Based Image Retrieval based on wavelet transform coefficients distribution", *Proc. IEEE Engineering in Medicine and Biology Society*, vol.1, pp.4532-4535, 2007.

[Lehman et al., 2006] T. Lehmann, T. Deselaers, H. Schubert, M. Güld, C. Thies, B. Fischer and K. Spitzer , "IRMA - a content-based approach to Image Retrieval in Medical Applications", *Proc. IRMA*, pp. 911-912, 2006.

[Leake, 1996] D. Leake, *CBR in context: the present and the future*, D.B. Leake, pp. 1-35, *AAAI/MIT* press, Menlo Park, 1996.

[Li and Ying, 2003] S.Li and M.Ying, "Region connection calculus: its models and composition table", *Artificial Intelligence*, vol. 145, no.1-2, pp. 121-146, 2003.

[Little and Hunter, 2004] S. Little and J. Hunter, "Rules-By-Example- A Novel Approach to Semantic Indexing and Querying of Images", *Proc.ISWC*, pp. 534-548, 2004.

[Liu et al., 2004] Y. Liu, N. Lazar, W. Rothfus, F. Dellaert, A. Moore, J. Schneider and T. Kanade, "Semantic – based Biomedical Image Indexing and Retrieval", *Trends and Advances in Content- Based Image and Video Retrieval*", Shapiro, Kriegel and Veltkamp eds., pp.1-20, 2004.

[Loménie et al., 2009] N.Loménie, L.Roux, D. Balensi, A. Tutac and D. Racoceanu, "MICO: The COgnitive Virtual Microscope project", *Cognitive Systems with Interactive Sensors (COGIS) symposium*, Paris**,** France, 16-18 Nov, 2009.

[Long et al., 2003] F. Long, H. Zhang and D. Feng, *Multimedia Information Retrieval and Management– Technological Fundamentals and Application,* D. Feng, W. C. Siu and Hongjing Zhang, Springer- Verlag, Germany, pp.1-26, 2003.

[Lopez and Perez, 2002] M. Fernandez Lopez and A. Gomez-Perez, "Overview and analysis of Methodologies for Building Ontologies", *The Knowledge Engineering Review*, vol 17, no. 2, pp. 129-156, 2002.

[Lovis et al., 2009] C. Lovis, T. Douglas, E. Pasche, P. Ruch, D. Colaert and K. Stroetmann, "DebugIT: building a European distributed clinical data mining network to foster the fight against microbial diseases", Stud Health Technol Inform, vol.148, pp. 50-59, 2009.

[Lukasiewicz, 2007] T. Lukasiewicz, "Probabilistic Description Logics for the Semantic Web", Tech. Rep.,Wien, TR 1843-06-05, 2007.

[Lundin et al., 2004] M. Lundin, J. Lundin, H. Helin and J. Isola, "A digital atlas of breast histopathology: an application of web based virtual microscopy", *Journal of Clinical Pathology*, vol. 57, pp. 1288-1291, 2004.

[Maiga and Williams, 2009] G.Maiga and D.Williams , "A Flexible Biomedical Ontology Selection Tool", *International Journal of Computing and ICT Research*, Special Issue Vol. 3, No. 1, pp. 53-66, 2009.

[Manning et al., 2008] C.Manning, P.Raghavan and H.Schutze, "Introduction to Information Retrieval", section 9- Relevance feedback and query expansion, pp.177-194, Cambridge Press, 2008.

[Maris, 2008] N.Maris, "A reasoner for querying temporal ontologies", Dissertation thesis, 2008.

[McGuiness and Harmelen, 2009] D.L.McGuiness and F.van Harmelen, "OWL Web Ontology Language Overview", http://www.w3.org/TR/owl-features/, 2009, last accessed July 2010.

[Mechouche et al., 2009] A. Mechouche, X. Morandi. C. Golbreich, and B. Gibaud, "A hybrid system using symbolic and numeric knowledge for the semantic annotation of sulco-gyral anatomy in brain MRI images", *IEEE Transactions on Medical Imaging,* vol. 28, no.8, pp. 1165-1178, Aug. 2009.

[Mejino et al., 2008] J. Mejino, D. Rubin, and J. Brinkley, "FMA-RadLex: An Application Ontology of Radiological Anatomy derived from the Foundational Model of Anatomy Reference Ontology," *AMIA Annual Symposium Proc.,* pp. 465–469, Nov. 2008.

[Mezaris et al., 2003] V.Mezaris, I. Kompatsiaris, and M. G. Strintzis, "An ontology approach to object-based image retrieval", IEEE International Conference on Image Processing, pp. 511-514, 2003.

[Mezaris et al., 2004] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Region-based Image Retrieval using an Object Ontology and Relevance Feedback", *EURASIP Journal on Applied Signal Processing*, Special Issue on *Object-Based and Semantic Image and Video Analysis*, vol. 2004, no. 6, pp. 886-901, June 2004.

[Menzel, 2007] Menzel C. "Reference Ontologies -- Application Ontologies: Either/Or or Both/And?", 2007, available from: http://bioontology.org/wiki/images/d/d9/MenzelOntology.pdf, accessed 2009

[Mulrane et al., 2008] L. Mulrane, E. Rexhepaj, S. Penney, J. Callanan, and W. Gallagher, "Automated image analysis in histopathology: a valuable tool in medical diagnostics," *Expert Review of Molecular Diagnostics*, vol. 8, no. 6, pp. 707–725, 2008.

[Müller et al., 2003] H. Müller, A. Rosset, J.P. Vallee and A. Geissbuhler, "Integrating content-based visual access methods into a medical case database", *Proc. of the Medical Informatics Europe Conference* (MIE), 2003.

[Müller, 2004] H. Müller, N. Michoux, D. Bandon and A. Geissbuhler, "A Review of Content-Based Image Retrieval System in Medical Applications- Clinical Benefits and Future Directions*", IJMI*, vol. 73, pp. 1-23, 2004.

[Nillson and Sollenborn, 2004] M. Nillson and M. Sollenborn, "Advancements and Trends in Medical Case-Based Reasoning: An overview of Systems and System Development", *Proc.FLAIRS*, pp.178-183, 2004.

[Obitko, 2007] M. Obitko, "OWL-DL Semantics, Introduction to ontologies and semantic web", http://www.obitko.com/tutorials/ontologies-semantic-web/owl-dl-semantics.html, 2007, last accessed July 2010.

[Ouagne et al., 2005] C. Le Bozec, E. Zapletal, M. Thieu and M.C. Jaulent, "Integration of Multiple Ontologies in Breast Cancer Pathology," in *Connecting Medical Informatics and Bio-Informatics: Proceedings of MIE2005 – The XIXth International Congress of the European Federation for Medical Informatics*, vol. 116/2005, pp. 641–646, 2005.

[Ouagne, 2009] D. Ouagne, "Open Medical Development Framework (OMDF) : méthodologie et environnement pour la conception et la mise en oeuvre de composants informatiques médicaux", PhD thesis, 2009.

[O'Connor and Das, 2009] M.O'Conor and A.Das, "SQWRL: A Query Language for OWL", Proc. of the 5th International Workshop on OWL: Experiences and Directions (OWLED), Rinke Hoekstra, Peter F. Patel-Schneider, editors, CEUR Workshop Proceedings vol. 529, pp. 1-8, 2009.

[Pal and Shiu, 2004] S. Pal and S. Shiu, *Foundations of Soft-Case-Based Reasoning*, Willey & Sons, pp. 4 -32, 2004.

[Paradiso et al., 2005] Italian Network for Quality Assurance of tumor biomarkers group (INQAT), "Quality control for histological grading in breast cancer: an Italian experience", Pathologica, vol.97, pp.1-6, 2005.

[Paradiso et al., 2009] A. Paradiso, I.Ellis, F. Zito, E. Marubini, S. Pizzamiglio, P.Verderio, "Short-and long-term effects of a training session on pathologists' performance: the INQAT experience for histological grading of breast cancer", Journal of Clinical Pathology, vol.62, no.3, pp. 279-281, 2009.

[Parsia et al., 2005] B.Parsia, E.Sirin, B.C.Grau, E.Ruckhaus and D.Hewlett, "Cautiously approaching SWRL", preprint submitted to Elsevier Science, 2005.

[Perner, 2001] P. Perner, "Why Case-Based Reasoning is Attractive for Image Interpretation", *Case-Based Reasoning Research and Development*, pp.27-43, 2001.

[Petushi et al., 2006] S. Petushi, F. Garcia., M.Haber, C. Katsinis, and A. Tozeren, "Large- Scale Computation on histology images reveal grade differentiating parameter for breast cancer", pp. 1-11, 2006.

[Pisanelli, 2004] D.Pisanelli, Ontologies http://www.openclinical.org/ontologies.html, 2004, last accessed July 2010.

[Quellec et al., 2008] G.Quellec, M.Lamard, L.Bekri, G.Cazuguel, C.Roux and B.Cochener, "Multi-modal medical case retrieval using decision trees", ITBM-RBM, vol.29, no.1, pp.35-43, 2008.

[Quellec et al., 2010a] G.Quellec, M.Lamard, L.Bekri, G.Cazuguel, C.Roux and B.Cochener, "Medical case retrieval from a committee of decision trees", IEEE Transaction on Information Technology on Biomedicine, vol.14, no.5, pp. 1227-1235, 2010.

[Quellec et al., 2010b] G.Quellec, M.Lamard, G.Cazuguel, B.Cochener and C.Roux, "Wavelet optimization for content-based image retrieval in medical databases", IEEE Transactions on Medical Image Analysis, vol.14, no. 2, pp.227-241, 2010.

[Quellec et al., 2010c] G.Quellec, M.Lamard, G.Cazuguel, B.Cochener and C.Roux, "Adaptive non-separable wavelet transform via lifting and its application to Content-Based Image Retrieval", IEEE Transactions on Image Processing, vol.19, no.1, pp.25-35, 2010.

[Racoceanu et al., 2009] D. Racoceanu, A.Tutac, W. Xiong, J-R. Dalle, C-H. Huang, L. Roux, W-K Leow, A. Veillard, J-H. Lim, T. Putti, and T. Ming, "A virtual microscope framework for breast cancer grading", *A*STAR CCO workshop in Computer Aided Diagnosis, Treatment and Prediction*, Biopolis, Singapore, January 2009.

[Rector et al., 2004] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang and C. Wroe, "OWL Pizzas: Practical Experiences of Teaching OWL-DL: common errors & common patterns", 2004.

[Richter, 2003] M. Richter, *Informal Introduction to Similarity-Based and Case Based Reasoning*, 2003.

[Rojo et al., 2006] M. Rojo, G. Garcia, C. Mateos, J. Garcia, and M. Vicente, "Critical comparison of 31 commercially available digital side systems in pathology," *International Journal of Surgical Pathology*, vol. 14, no. 4, 2006.

[Roux et al., 2009a] L. Roux, A. Tutac, N. Lomenie, D. Balensi, A. Veillard, D. Racoceanu, W.K. Leow, J. Klossa, and T.C. Putti, "A cognitive virtual microscopic framework for knowledge-based exploration of large microscopic images in breast

cancer histopathology", in *Proc.* IEEE Engineering in Medicine and Biology Society, pp. 3697-3702, SUA, 2009.

[Roux et al., 2009b] L.Roux, A.Tutac, A. Veillard , J. Dalle , D. Racoceanu, N. Lomenie and J. Klossa, "A cognitive approach to microscopy analysis applied to automatic breast cancer grading" , Virchows Archiv The European Journal of Pathology, Springer-Verlag Berlin Heidelberg, H.Höfler ed, no. 428, vol. 455, supplement 1: S1-S482, DOI 10.1007/s00428-009-0805-z, pp.S34, ISSN : 0945-6317 (Print) 1432-2307 (Online), *22nd European Congress of Pathology*, Florence, Italy, 4-9 Sept 2009.

[Samuel et al., 2008] K. Samuel, L. Obrst, S. Stoutenberg, K. Fox, P. Franklin, A. Johnson, K. Laskey, D.Nichols, S. Lopez and J. Peterson "Translating OWL and Semantic Web Rules in Prolog: Moving toward Description Logic Programs", *Theory and Practice of Logic Programming*, vol.8, nr. 3, pp. 301-322, 2008.

[Sciacio et al., 2002] E.di Sciascio, F.Donini and M.Mongiello, "Structured Knowledge representation for image retrieval", *Journal of Artificiall Intelligence Research*, vol. 16, pp. 209-257, 2002.

[Schmidt and Gierl, 2001] R. Schmidt and L. Gierl, "Medical Case-Based Reasoning Systems: Experiences with Architectures for Prototypical Cases", *Proc.MEDINFO*, pp.518-522, 2001.

[Schmidt et al., 2003] Schmidt R., Vorobieva O. and Gierl L., Case-based Adaptation Problems in Medicine, *Proc. WM*, vol. 67, pp.1-9, 2003.

[Schulz et al., 2005] S. Schulz, P. Daumke, B. Smith, and U. Hahn, "How to distinguish between parthood and location in bio-ontologies", *AMIA Annual Symposium Proc.,* pp.669- 673, 2005.

[Seidenberg and Rector, 2006] J. Seidenberg , A. Rector, "Web ontology segmentation: analysis, classification and use", *Proceedings of the 15th international conference on World Wide Web*, pp. 13-22, NY, SUA, ACM Press, 2006.

[Shadbolt et al., 2006] N. Shadbolt, W. Hall, and T.Berners-Lee, "The Semantic Web Revisited, *IEEE Intelligent Systems*, pp. 96-101, 2006

[Shanmugaratnam, 2007] K. Shanmugaratnam, "Happenings in histopathology — a post-World War II perspective", *Annals Academy of Medicine, Singapore*, Tech. Rep., vol 36, pp. 691-697, 2007.

[Sirin et al., 2005] E. Sirin., B. Parsia, B.Grau, A. Kalyanpur and Y. Katz, "Pellet: A practical OWL-DL reasoner", pp. 1-26, 2005.

[Smeulders et al., 2000] A. Smeulders, M. Worring, S. Santini, A. Gupta and R.Jain, "Content -Based at the End of the Early Years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22 no.12, pp.1349-1380, 2000.

[Smith, 2004] B. Smith, "Beyond Concepts: Ontology as Reality Representation", *Proc.FOIS*, pp. 1- 12 2004.

[Soenksen, 2005] D. Soenksen, "A fully integrated virtual microscopy system for analysis and discovery," *Virtual Microscopy and Virtual Slides in Teaching, Diagnosis and Research*, pp. 35–47, 2005.

[Steichen et al., 2006] O. Steichen, C.D- Le Bozec, M. Thieu, E. Zapletal, and M.C. Jaulent, "Computation of semantic similarity within an ontology of breast pathology to assist inter-observer consensus," *Computers in Biology and Medicine*, vol. 36, no. 7, pp. 768–788, 2006.

[Stocker and Sirin, 2009] M.Stocker and E.Sirin, "PelletSpatial: A Hybrid RCC-8 and RDF/OWL Reasoning and Query Engine", *Proc. OWL: Experiences and Directions* (OWLED), 6th International Workshop, vol. 529, 2009.

[Stroińska and Hitchcock, 2002] M. Stroińska and D. Hitchcock, "On the Concept of Following Logically", *History and Philosophy of Logic*, vol.23, pp.155-196, 2002.

[Styrman, 2005] A. Styrman, Ontology-based image annotation and retrieval, Master thesis, 2005.

[Tadrat et al., 2007] J.Tadrat, V. Boonjing and P. Pattaraintakorn, "A Framework for using Rough Sets and Formal Concept Analysis in Case Based Reasoning", *Proc.IRI*, pp.227-232, 2007.

[Tang et al., 2003] H.Tang, R.Hanka, and H.Ip, "Histological Image Retrieval Based on Semantic Content Analysis", *IEEE Transaction on Information Technology in Medicine*, vol. 7, no.1, pp.26-36, 2003.

[Tartir et al., 2005] S.Tartir, B.Arpinar, M.Moore, A.Seth and B.Aleman-Meza, "OntoQA:Metric-Based Ontology Quality Analysis", in *IEEE Workshop on Knowledge Acquisition fromDistributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, USA, IEEE Computer Society, pp. 45 -53, 2005.

[Traina et al., 2005] J.C. Traina, A.J.Traina, M. Araujo J. Bueno, F. Chino, H. Razente and P.Azevedo-Marques, "Using an image-extended relational database to support content-based image retrieval in a PACS", *Computer Methods and Programs in Biomedicine*, vol. 80 no.1, pp. 71-83, 2005.

[Tutac et al., 2008a] A.E. Tutac, D. Racoceanu, T. Putti, W. Xiong, W.K. Leow, and V. Cretu, "Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images", BioMedical Engineering and Informatics: New Development and the Future, *in Proc. BMEI*, Yonghong Peng & Yufeng Zhang, Ed, vol.2, pp. 107-112, 2008.

[Tutac et al., 2008b] A.E. Tutac, D. Racoceanu, W.K. Leow, J.R. Dalle, T. Putti, W. Xiong, and V. Cretu, "Translational Approach for Semi-Automatic Breast Cancer Grading using a Knowledge-Guided Semantic Indexing of Histopathology Images", in *Proc. MIAAB*, 2008.

[Tutac et al., 2009a] A.Tutac, D.Racoceanu, W.Leow, H. Muller, T.Putti and V.Cretu, "Towards translational incremental similarity-based reasoning in breast cancer grading", in *Proc. SPIE Medical Imaging: Computer Aided Diagnosis*, Nico Karssemeijer, Maryellen L.Giger eds, vol.7260, 72603C, pp.1-12, Feb 2009, in revue of Progress of biomedical optics and imaging, vol.10 (2), no.36, ISSN: 1605-7422, SUA, 2009.

[Tutac et al., 2009b] A.Tutac, D. Racoceanu, N. Loménie, L. Roux, T. C. Putti, and V. Cretu, "Breast Cancer Grading Knowledge Modeling and Reasoning for Cognitive Virtual Microscopy", *National Institutes of Health NIH Inter-Institute Workshop on Optical Diagnostic and Biophotonic Methods from Bench to Bedside*, Bethesda, USA, 1-2 Oct 2009.

[Tutac et al., 2009c] A. Tutac, D. Racoceanu, N. Loménie, L. Roux, D. Balensi and T. Putti, "Knowledge Representation and Reasoning for Breast Cancer Grading in Cognitive Virtual Microscope Framework", *A*STAR Scientific Conference* 2009, Biopolis, Singapore, 28-29 Oct, 2009.

[Tutac et al., 2009d] A. Tutac, D.Racoceanu, N.Loménie , W.K.Leow., L.Roux, V.I.Cretu and T. Putti , "Knowledge Modeling of Breast Cancer Grading using OWL-DL formalism", Virchows Archiv The European Journal of Pathology, Springer-Verlag Berlin Heidelberg, H. Höfler ed, no. 428 vol. 455, no.1: S1-S482, DOI 10.1007/s00428-009-0805-z, pp. S36, ISSN : 0945-6317 (Print) 1432-2307 (Online), *22nd European Congress of Pathology*, Florence, Italy, 4-9 Sept 2009.

[Tutac et al., 2010a] A. Tutac, V. Cretu and D. Racoceanu, "Spatial representation for Breast Cancer Grading Ontology", *Proc. IEEE International Joint Conferences on Computational Cybernetics and Technical Informatics ICCC-CONTI* , pp. 89-94, Timisoara, Romania, 27-29 May, 2010.

[Tutac et al., 2010b] A.Tutac, V. Cretu and D. Racoceanu, "A Spatial representation and Reasoning Approach for Breast Cancer Grading Ontology", Scientific Bulletin of Politehnica University of Timisoara, Transactions on Automatic Control and Computer Science, vol.55, no.69, issue 3, pp. 123-133, ISSN: 1224-600X, September 2010.

[Uschold and Grüninger, 1996] M. Uschold and M. Grüninger, "Ontologies: Principles, Methods and Applications," *Knowledge Engineering Review*, vol. 11, no. 2, pp. 93–155, 1996.

[Vacura et al., 2008] M, Vacura, V.Svatek, C. Saathoff, T. Franz and R. Troncy, "Describing low-level image features using the COMM ontology", *Proc.International Conference on Image Processing*, pp. 49-52, 2008.

[Varzi, 2009] A.Varzi, "Mereology", Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/mereology/, accessed 2010

[Vasconcelos, 2007] N. Vasconcelos, "From Pixels to Semantic Spaces: Advances in Content-Based Image Retrieval", *Computer,* vol.40 no.7, pp. 20 -26, 2007.

[Veillard et al., 2010] A. Veillard, N. Lomenie and D. Racoceanu, "An Exploration Scheme for Large Images: application to Breast Cancer Grading", *International Conference on Pattern Recognition*, ICPR'2010, Istanbul, Turkey, August 23-26, 2010.

[Watson and Marir, 1994] I. Watson and F. Marir, Case-Based Reasoning: A Review. The Knowledge Engineering Review, vol. 9 no. 4, pp. 355-381, 1994.

[Watson, 1999] I. Watson, "Case based reasoning is a methodology not a technology", *Knowledge Based Systems*, vol. 12 no. 5-6, pp. 303- 308, 1999.

[Wang and Fartha, 2005] Z. Wang and A. Farha, "An Efficient Ontology Comparison Tool for Semantic Web Applications", *IEEE WIC ACM International Conference on Web Intelligence*, 2005.

[Wang and Parsia, 2008] T. Wang, and B. Parsia, "Ontology performance profiling and model examination: first steps", in *Semantic Web,* vol. 4825, Springer Berlin/Heildelberg, eds, pp.595-608, 2008.

[Wang et al., 2006] H. Wang, S. Liu, and L-T. Chia, "Does Ontology Help in Image Retrieval? — A Comparison between Keyword, Text Ontology and Multi-Modality Ontology Approaches", *14th Annual ACM International Conference on Multimedia*, ACM Press, pp. 109-112, 2006.

[Weinstein et al., 2005] R. Weinstein, M. Descour, C. Liang, L. Richter, W. Russum, J. Goodall, P. Zhou, A. Olzak, and P. Bartels, "Reinvention of light microscopy: Array microscopy and ultrarapid scanned virtual slides for diagnostic pathology and medical education," *Virtual Microscopy and Virtual Slide in Teaching, Diagnosis and Research*, pp. 9–34, 2005.

[Weinstein et al., 2007] R.S. Weinstein, A.M. López, G.P. Barker, E.A. Krupinski, M.R. Descour, K.M. Scott, L.C. Richter, S.J. Beinar, M.J. Holcomb, P.H. Bartels, R.A. McNeely, and A.K. Bhattacharyya, "The innovative bundling of teleradiology, telepathology and teleoncology services," *IBM Systems Journal*, vol. 46, pp. 69–84, 2007.

[Weinstein, 2007] R. S. Weinstein, "View master — an expert eyes digital pathology's future," in *Futurescape of Pathology CAP Foundation conference*, 2007.

[w3reg] Region connection calculus, http://en.wikipedia.org/wiki/Region_connection_calculus, accessed 2010

[Zhang et al., 2006] S. Zhang, O.Bodenreider and C.Golbreich, "Experiences in reasoning with the Foundational Model of Anatomy in OWL-DL", *Pacific Symposium on Biocomputing*, vol.11, pp.200-211, 2006.

[Zhao and Groski, 2001] R. Zhao and W. Groski, "Bridging the Semantic Gap in Image Retrieval", *Distributed Multimedia Databases: Techniques and Applications*, T. K. Shih, Idea Group, pp. 14-36, 2001.

Research Activity and Publications

PhD internships

1. March – June 2009, IPAL (CNRS UMI 2955, A*STAR/I^2R, NUS, UJF) & NUH Singapore

Research subject: "Interpretive Hybrid Probabilistic Reasoning in Microscopic Medical Image-Based Prognosis and Research/ Use of Histopathology Images in Breast Cancer Grading Framework**"**
Supervisor from CNRS/UFC: A/Prof. Daniel RACOCEANU HDR
Supervisor from UPT: Prof. Vladimir-Ioan CRETU PhD
Collaborator from NUH, Pathology Department: A/Prof. Thomas Choudary Putti MD

2. March - June 2008, IPAL (CNRS UMI 2955, A*STAR/I^2R, NUS, UJF) & NUH Singapore

Research subject: "CBIR versus CBR towards hybrid reasoning in Breast Cancer Grading"
(MMedWeb project)
Supervisor from CNRS/UFC& NUS: A/Prof. Daniel RACOCEANU HDR
Supervisor from UPT: Prof. Vladimir-Ioan CRETU PhD
Collaborator from NUH, Pathology Department: A/Prof. Thomas Choudary Putti MD

3. March - July 2007, IPAL (CNRS UMI 2955, A*STAR/I^2R, NUS, UJF) & NUH Singapore

Research subject: "Knowledge – Guided Semantic Indexing of Breast Cancer Histopathology Images"
(**ONCO-MEDIA project**)
Supervisor from CNRS/UFC& NUS: A/Prof. Daniel RACOCEANU HDR,
Supervisor from UPT: Prof. Vladimir -Ioan CRETU PhD
Collaborator from NUH, Pathology Department: A/Prof. Thomas Choudary Putti MD

Research grants & patents

1. National University Research Council CNCSIS, Romania, research grant PNII, "Micro-medical Image Processing", type TD, contract no. 123/17.09.2008, Director: drd.ing. Tutac Adina.

2. Patent - software declaration inventoried as *DI 2944-01* by the CNRS for the *UMI 2955*. Registered by the CNRS, "HISTOGRAD – a virtual

microscope for breast cancer grading" **,** Daniel Racoceanu, Adina Tutac, Xiong Wei, Jean-Romain Dalle, Chao-Hui Huang, Ludovic Roux, Wee-Kheng Leow.

Journal

1. A.Tutac, V. Cretu and D. Racoceanu, "A Spatial representation and Reasoning Approach for Breast Cancer Grading Ontology", Scientific Bulletin of Politehnica University of Timisoara, Transactions on Automatic Control and Computer Science, vol.55, no.69, issue 3, pp. 123-133, ISSN: 1224-600X, September 2010.

International Conferences

1. **A.Tutac**, V. Cretu and D. Racoceanu, "Spatial representation for Breast Cancer Grading Ontology", Proc. IEEE International Joint Conferences on Computational Cybernetics and Technical Informatics ICCC-CONTI, pp. 89-94, ISBN: 978-1-4244-7431-8, Timisoara, Romania, 27-29 May, 2010, (IEEE indexed)

2. L. Roux, **A. Tutac**, N. Lomenie, D. Balensi, A. Veillard, D. Racoceanu, W.K. Leow, J. Klossa, T.C. Putti, "A cognitive virtual microscopic framework for knowledge-based exploration of large microscopic images in breast cancer histopathology", in Proc. IEEE Engineering in Medicine and Biology Society, pp. 3697-3702, Minneapolis, SUA, 2-6 Sept 2009, (IEEE indexed)

3. Roux L., **Tutac A**., Veillard A., Dalle J., Racoceanu D., Lomenie N., Klossa J, "A cognitive approach to microscopy analysis applied to automatic breast cancer grading" , Virchows Archiv The European Journal of Pathology, Springer-Verlag Berlin Heidelberg, H.Höfler ed, no. 428, vol. 455, no 1: S1-S482, DOI 10.1007/s00428-009-0805-z, pp.34-35, ISSN : 0945-6317 (Print) 1432-2307 (Online), 22nd European Congress of Pathology, Florence, Italy, Sept 2009, (ISI indexed, http://www.springerlink.com/content/l7w61687q067g047/)

4. **Tutac Eunice A**., Racoceanu D., Lomenie N., Leow Kheng W., Roux L., Cretu I.V., Putti T, "Knowledge Modeling of Breast Cancer Grading using OWL-DL formalism", Virchows Archiv The European Journal of Pathology**,** Springer-Verlag Berlin Heidelberg, H. Höfler ed, no. 428 vol. 455, no.1: S1-S482, DOI 10.1007/s00428-009-0805-z, pp. 36, ISSN : 0945-6317 (Print) 1432-2307 (Online), 22nd European Congress of Pathology, Florence, Italy, Sept 2009, (SpringerLink indexed, http://www.springerlink.com/content/l7w61687q067g047/)

5. **A.E.Tutac**, D. Racoceanu, W. Leow, H. Müller,, T. Putti, V. Cretu, "Toward translational incremental similarity-based reasoning in breast cancer grading" in Medical Imaging 2009: Computer-Aided Diagnosis, Proc. SPIE (SPIE, Bellingham, WA 2009), Vol. 7260, 72603C (2009); Nico Karssemeijer, Maryellen L.Giger eds, ISBN : 978-0-8194-7511-4, pp. 1-12,

Orlando, Florida, SUA, 7-12 February 2009 (IEEE, CAT-INIST indexed http://cat.inist.fr/?aModele=afficheN&cpsidt=21807004)

6. J.Dalle, D. Racoceanu, W.K Leow**, A.E. Tutac**, T.Putti, "Automatic Breast Cancer Grading of Histopathological Images", Proc of 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE EMBS, ISBN 978-1-4244-1814-5, ISSN 1557-170X, pp.3052-3055, Vancouver, Canada, 2008 (ISI, IEEE, COMPEDEX indexed, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4649847)

7. **A.E. Tutac**, D. Racoceanu, T. Putti, W. Xiong, W.K. Leow and V. Cretu, "Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images", BioMedical Engineering and Informatics: New Development and the Future, Proc. BMEI, ed.Yonghong Peng & Yufeng Zhang, ISBN 978-0-7695-3118-2,vol.2, pp. 107-112, China, 2008, (ISI & IEEE indexed http://portal.acm.org/citation.cfm?id=1372330)

Symposiums & Workshops

1. N.Loménie, L.Roux, D. Balensi, **A. Tutac**, D. Racoceanu, "MICO: The COgnitive Virtual Microscope project", Cognitive Systems with Interactive Sensors (COGIS) symposium, Paris**,** France, 16-18 Nov, 2009.

2. **Adina Tutac**, Daniel Racoceanu, Nicolas Loménie, Ludovic Roux, Didier Balensi and Thomas Putti, "Knowledge Representation and Reasoning for Breast Cancer Grading in Cognitive Virtual Microscope Framework", A*STAR Scientific Conference 2009, Biopolis, Singapore, 28-29 Oct, 2009.

3. **Adina Tutac**, Daniel Racoceanu, Nicolas Loménie, Ludovic Roux, Thomas C. Putti, Vladimir Cretu, "Breast Cancer Grading Knowledge Modeling and Reasoning for Cognitive Virtual Microscopy", National Institutes of Health NIH Inter-Institute Workshop on Optical Diagnostic and Biophotonic Methods from Bench to Bedside, Bethesda, USA, 1- 2 Oct 2009

4. D. Racoceanu, **A.Tutac**, W. Xiong, J-R. Dalle, C-H. Huang, L. Roux, W-K Leow, A. Veillard, J-H. Lim, T. Putti, T. Ming, "A virtual microscope framework for breast cancer grading", A*STAR CCO workshop in Computer Aided Diagnosis, Treatment and Prediction, Biopolis, Singapore, 15 January 2009.

5. **A.E. Tutac**, D. Racoceanu, W.K. Leow, J.R. Dalle, T. Putti, W. Xiong and V. Cretu, "Translational Approach for Semi-Automatic Breast Cancer Grading Using a Knowledge-Guided Semantic Indexing of Histopathology Images", 3rd Microscopic Image Analysis with Application in Biology MIAAB Workshop, in conj. with MICAAI, 11th International Conference on Medical Image Computing and Computer Assisted Intervention, SUA, 6-10 Sept, 2008, COMPEDEX, INSPEC indexed.

PhD/Technical reports

1. **A.Tutac**, "Content-Based Image Retrieval versus Case-Based Reasoning towards Hybrid reasoning in Breast Cancer grading", PhD report #1, Timisoara, February 2009

2. **A. Tutac,** "Formal Representation and Reasoning for Breast Cancer Grading", PhD report #2, April 2010

3. J-R.Dalle, **A.Tutac**, "Breast Cancer Report", IPAL, TR-dalle20071217ipal_mmedweb, Singapore, Dec 2007

4. **A.Tutac**, "Histological Grading of Breast Cancer", IPAL, TR-tutac20070407ipal_oncomedia, Singapore, May, 2007

Annexe

REPRÉSENTATION ET RAISONNEMENT FORMELS POUR LE PRONOSTIC BASÉ SUR L'IMAGERIE MÉDICALE MICROSCOPIQUE. APPLICATION À LA GRADUATION DU CANCER DU SEIN.

A1. Résumé Etendu

La présente étude est le résultat de recherches concernant une approche de représentation formelle qualitative de la connaissance médicale, avec l'aide d'ontologies et de mécanismes d'inférence, pour l'assistance automatique du pronostic du cancer du sein. Nos travaux visent à développer une ontologie du cancer du sein appelé BCGO, permettant d'obtenir une efficacité et une reproductibilité de cet examen médical, par rapport à la procédure manuelle. L'objectif est d'assurer une terminologie formelle normalisée, pour bénéficier d'une expressivité élevée et de puissance de calcul, et pour permettre l'intégration de l'ontologie dans une plateforme virtuelle cognitive (comme une assistant virtuel pour le pronostic) ainsi que des recherches ultérieures concernant l'extension de la représentation formelle pour des problèmes visuels.

Le document est structuré en neuf chapitres :

Le **Chapitre 1** définit le contexte et les objectifs de la thèse. La représentation formelle des connaissances dans divers domaines est une composante essentielle des systèmes perceptifs et cognitifs. Une direction d'intérêt pour le présent document est la représentation d'espace, en raison du lien étroit avec l'aspect perceptif et cognitif. Ces représentations peuvent être classifiées en: quantitatives (numériques) et qualitatives (symboliques), selon le contexte et le domaine de la représentation. Dans les applications médicales, une représentation qualitative est confrontée à la « métaphore d'aquarium », dans laquelle - au moment du diagnostic ou du pronostic - les informations ne sont pas toutes disponibles (par exemple, le radiologue n'a pas d'informations sur la pathologie), ou le diagnostic/pronostic dépend de l'interprétation donnée par le praticien, de la perception et l'expérience personnelle de celui-ci. Enfin, les images qui sont étroitement liées à l'espace, jouent un rôle de plus en plus important dans le diagnostic et le pronostic.

La graduation du cancer du sein (BCG) est un bon exemple, dans ce contexte. Elle est actuellement considérée comme essentielle pour le pronostic dans la pratique de pathologie moderne. Le système de classement Nottingham (NGS – Nottingham Grading System) est le système standard utilisé par les pathologistes. Ce système repose sur l'analyse manuelle des tissus microscopiques dans la forme de lame ou de cadres (champs optiques microscopiques), en appliquant trois critères d'analyse: l'identification des formations tubulaires, pleomorphism nucléaire et le nombre de mitoses.

La recherche et l'expérience médicale a montré que cette analyse est fortement influencée par l'expérience de chaque pathologiste. Par ailleurs, le manque de représentation sémantique formelle standardisée pour aider l'indexation et la classification de la terminologie, ainsi que l'utilisation d'un mécanisme d'inférence pour assister le pronostic représentent des problématiques clé du domaine.

En conséquence, les objectifs de cette thèse sont :

- une étude de deux approches différentes de l'indexation basée sur les caractéristiques d'imagerie (leur contenu) et de ceux obtenus à partir de la description des cas : Content-Based Image Retrieval (CBIR) et Case-Based Reasoning (CBR).
- une nouvelle méthode pour lever les incohérences est l'accord des pathologistes concernant le développement d'un modèle formel qualitatif (appelé Breast Cancer Grading Ontology - BCGO) comme une ontologie d'application.
- une théorie de la représentation spatiale pour les relations/concepts BCG liés à l'espace exploré, l'intégration des concepts mereo-topologique, métrique, géométrique et d'échelle.
- une méthode d'évaluation qualitative avec une validation médicale pour BCGO
- l'intégration du modèle formel BCGO dans une plateforme de microscopie virtuelle MICO, comme support pour l'annotation, l'exploration visuelle et l'extraction de concepts sémantiques liés aux caractéristiques de l'image et donc représentant un microscope virtuel élevé au niveau cognitif.

Le **Chapitre 2** est destiné à faire une incursion dans les approches conceptuelles liées à la représentation formelle et aux mécanismes d'inférence, les représentations d'images existantes et les représentations au niveau sémantique. Deux stratégies ont été identifiées lors de la représentation d'image: Content-Based Image Retrieval (CBIR) et Case Based Reasoning (CBR) et trois techniques ont été étudiées en détail: l'indexation (*indexing*), l'extraction (*retrieval*) et le raffinage (*refining*), en soulignant leurs avantages et leurs inconvénients à chaque niveau. Sur la base de leur analyse du point de vue technologique et méthodologique, il est soutenu que les deux paradigmes peuvent être considérés comme une méthodologie et il est proposée une combinaison CBR-CBIR comme une nouvelle méthodologie qui utilise les avantages de chaque approche en partie, pour une application dans le domaine médical. Au niveau sémantique, les langages formels Description-Logics (DL), Web Ontology Language (OWL) et Semantic Web Rule Language (SWRL) ont été analysés et aussi les types d'ontologies, afin d'identifier une représentation plus efficace pour la graduation du cancer du sein. La dernière section de ce chapitre présente un état de l'art concernant la représentation spatiale, en mettant l'accent sur la représentation qualitative et les diverses théories qu'ont été proposées: mereologie, topologie, mereo-topologie et la théorie de la représentation de la distance et de l'orientation.

Le **Chapitre 3** est consacré à l'analyse de l'application du CBIR et du CBR en médecine : la représentation de l'image et le niveau de représentation sémantique dans lequel sont présentes les ontologies bio-médicales, compilées à partir des cadres et des réseaux sémantiques, sont comparées à ceux qui utilisent le formalisme logique. Les raisons pour lesquelles une représentation formelle ontologique résolve le fossé sémantique (*semantic gap*) et le fossé contextuel (*context gap*) générique défini come le fossé de contenu (*content gap*) sont aussi présentées. Notre étude prépare donc des fondations pour le développement d'un nouveau modèle ontologique pour le BCG.

L'analyse CBIR-CBR a conduit à l'affirmation de l'un des principes de l'approche proposée dans cette thèse: le raisonnement est commun dans les deux CBIR et le CBR. Le raisonnement à CBIR est basé sur l'image tandis que le CBR est un raisonnement fondé sur des informations structurées sur les cas, essentiellement textuels. En termes de logique formelle, le raisonnement est le mécanisme d'inférence. Ainsi, le mécanisme d'inférence est essentiel dans le CBIR et CBR, et il représente un soutien pour l'indexation sémantique et la recherche de concepts sémantiques liés a les caractéristiques des objets de l'image. Un autre principe qui découle de l'analyse des ontologies bio-médicale est que les ontologies, avec un mécanisme d'inférence DL sont la clé d'une représentation structurée et calculable des connaissances du domaine d'intérêt. Aussi, parce que les images histopathologiques jouent un rôle important dans le pronostic, l'accent est mis sur la nécessité d'utiliser une représentation spatiale formelle et une théorie de l'espace qui peut éliminer les incohérences et les ambiguïtés de la représentation. Sur la base de ces principes, nous allons construire le modèle formel présenté au chapitre 4 et la théorie spatiale du chapitre 5.

Le **Chapitre 4** présente le BCG, le système de classement standard NGS utilisé par les pathologistes, et met en évidence les principaux problèmes posés par la procédure, notamment du à une analyse qui nécessite connaissances, attention et temps (l'analyse est effectué sur des images microscopiques de très grandes dimensions, une lame contenant plus de 4000 cadres qui sont numérisés et étudiés sous le microscope). Ensuite le modèle ontologique proposé, dédié à la représentation des concepts médicaux liés a BCG. Deux modalités sont proposées*: image-driven* et *semantic-driven*. L'*image-driven* model est construit en deux étapes: *knwledge acqusition et knowledge translation.* Dans une premiere étape, les connaissances sont formalisées par des pathologistes (*subjective knowledge)* et sont extraites de la procedure NGS (*objective knowledge). Knowledge translation* est fondée sur un traducteur MK-CV (Medical Knowledge-Computer Vision). Ce modèle présente certaines limitations toutefois: il n'a pas la capacité de calcul de la logique formelle et les algorithmes de traitement d'image peuvent travailler de façon indépendante, ne possédant pas nécessairement de niveau sémantique (comme soutien ontologique). D'autre part, l'expressivité de la représentation n'est pas aussi élevée. Il est donc proposé le deuxième modèle, qui est l'ontologie d'application BCGO. Il vise à la modélisation qualitative des concepts de type perdurant (objets) avec le soutien d'une théorie formelle de l'espace. Contrairement au premier modèle, la conception de l'ontologie vise ici trois étapes:

- *knowledge acquisition* – les connaissances médicales sont acquises, d'une part par la segmentation de l'ontologie de référence National Cancer Institute (NCI) contenant des concepts liés au cancer du sein, et d'autre part, par le NGS contenant des concepts spécifiques pour la graduation. Il y a ici une approche différente, plus complète que le processus d'acquisition du premier modèle.
- *knowledge translation* - est basé sur un module structurel qui utilise la logique OWL-DL et un module de règles, construit en utilisant le langage SWRL. L'avantage de cette approche est d'obtenir une expression forte et en même temps, la puissance de décidabilité spécifique à la logique formelle, contrairement à d'autres ontologies biomédicales qui n'ont pas de mécanisme d'inférence ou de règles.
- *knowledge refining* – est basé sur les retour d'expérience (feedback) médical et sur le raisonneur DL Pellet, et atteindre ainsi le modèle déduit de l'ontologie.

La théorie formelle et le mécanisme d'inférence spatiale sont présentés dans le **Chapitre 5**. La théorie formelle proposée traite des concepts mereo-topologique, métrique, géométrique et d'échelle, pour aider à la cohérence de la représentation spatiale. Sont ainsi définies des relations mereo-topologiques comme *SurroundedBy,* à partir de la définition, les axiomes et les théorèmes de la relation *LocatedIn* proposée pour les ontologies biomédicales FMA et GALEN. Sont également définies des relations métriques comme *CloseTo.* Dans ce cas, une autre méthode de représentation qualitative est utilisée, à savoir, la Region Connection Calculus (RCC-8) et le tableau de composition. En utilisant la théorie spatiale, d*es ambiguïtés et des incohérences de la représentation peuvent être éliminés,* une question également illustré dans ce chapitre. Le mécanisme d'inférence est réalisé de deux façons: manuellement ou automatiquement. Le mécanisme manuel consiste à appliquer les axiomes et les théorèmes de la théorie de la représentation aux connaissances du BCG. Ce mécanisme examine en particulier leur application à partir du niveau *individuel* (instances) aux *classes*. L'algorithme d'inférence automatique proposée est basé sur l'algorithme tableau mis en ouvre dans le mécanisme d'inférence Pellet DL, l'outil utilisé dans le présent document pour le procès de classification automatique est la vérification de cohérence de l'ontologie. Le fonctionnement de cet algorithme dans BCGO est présenté et les résultats du raisonnement logique sont analysés.

Le **Chapitre 6** contient une description détaillée de l'implémentation de BCGO à partir du modèle proposé dans le chapitre 4, en utilisant l'environnement de programmation Protege-OWL (basé sur le langage de programmation Java) et le raisonneur Pellet (langage formel basé sur DL). Au cours de l'implémentation, sont discutées les principales caractéristiques de OWL, DL et de SWRL, la description des classes, le propriétés, les instances et les règles. L'implémentation rencontre l'étape de traduction (*knowledge translation)* et de raffinement (*knowledge refining)* de la base des connaissances. Le procès commence par la division de la base DL knowledge-base en *TBox* et *ABox*. *TBox* contient les axiomes, les définitions des concepts et les contraintes, précisant qu'en termes d'OWL ils représentent les classes, et la ABox contient des assertions de concepts et de propriétés (les relations) que représente les individus en termes d'OWL. Quatre catégories de classes sont définies: *AnatomicalEntity, ConceptualEntity, MicroscopicalEntity* et *SpatialEntity.* BCGO contient un nombre total de 129 classes, 169 individus et 86 relations. Il est montré ensuite, comment les classes OWL-DL sont construites - par exemple les classes définies et les classes primitives, classes disjointes - et il est illustré le principe de l'Open World Assomption (OWA) qui fonctionne par OWL, et qui implique un axiome de clôture dans la représentation de la classe. Sont par ailleurs discutés les propriétés des objets, *object properties et datatype properties* et mettent en œuvre des procédures qui assurent la plus grande expressivité et de décidabilité base sur le type particulier de propriété. Le module de règles SWRL peut fonctionner indépendamment ou en lien avec OWL, certaines règles pouvant être traduites en syntaxe OWL sans utiliser SWRL (appelées *syntactic sugar rules*). D'autre part, bien que les règles exprimées dans un langage SWRL complètent la représentation, se pose le problème d'indécidabilité. La solution proposée est d'utiliser les règles *SWRL DL safe*.

Le **Chapitre 7** présentes l'évaluation de BCGO en termes de qualité et des contraintes syntaxiques DL. Les indicateurs qualitatifs utilisés sont : la granularité, le degré de représentation et le degré d'intégration de BCGO ontologie dans autres ontologies. La granularité de BCGO représentation est analysée en fonction de la richesse des relations (*relationship richness*) et de la richesse des attributs (*atribute richness*). Le degré de réflexion du domaine médical de graduation du cancer du sein (BCG) représenté en BCGO est évalué en termes de richesse de la classe (*class richnesss*) et de population moyenne (*average population*). Le degré d'intégration de

l'ontologie BCGO dans autres l'ontologies examine les liens ou relations avec d'autres concepts d'ontologies avec qui BCGO a des éléments communs. Sont ainsi discutés des aspects liés aux contraintes syntaxiques DL, comme les restrictions cardinales sur les propriétés transitives ou les individuels dans l'énumération. La validation de l'ontologie est basée sur l'extration sémantique (*semantic retrieval) qui* se compose de requêtes sémantiques (*semantic queries)* de type RDF/SQWRL et aussi du retour d'expérience medical (*medical feedback*) et sur l'intégration de l'ontologie dans la plateforme Open Biomédical Ontologies. Il est conclut que l'ontologie met largement en œuvre en miroir le domaine médical et peut être intégré dans une prochaine étape dans une ontologie de référence.

Le **Chapitre 8** présente une étude de cas: l'intégration de l'ontologie BCGO dans une plateforme de microscopie virtuelle MICO (construite par l'équipe du laboratoire CNRS IPAL), où l'ontologie est essentielle pour l'aspect cognitif. L'approche de ce chapitre est basée sur la définition de la microscopie virtuelle et met en évidence l'importance de la microscopie virtuelle dans le contexte de la graduation du cancer du sein, suivie par la présentation des principes et des caractéristiques de la microscopie cognitive. Il est examiné ensuite que la plateforme MICO peut être considérée comme une plateforme microscopique cognitive, grâce à l'intégration et à l'utilisation de cette ontologie et des techniques de raisonnement associées.

La plateforme de MICO est divisée en quatre modules: *acquisition d'images, identification des Region d'Interet invasives (ROI), la graduation des ROI* et *le support de validation base sur les ontologies*. Dans le premier module mis en ouvre par l'équipe IPAL, il est effectué l'acquisition des images qui sont ensuite converties dans le format Digital Imaging and Communications in Medicine (DICOM) – en suivant le récent supplément 122 de ce standard (dédié a l'histopathologie). Il est ensuite appliqué un algorithme de *stitch and bleding* sur les frames pour former la lame virtuelle complète appelée WSI (Whole Slide Imaging). Dans le module de détection de *ROI*, sont appliqués des algorithmes de traitement d'image au niveau multi-échelle, permettant de graduer le pléomorphisme nucléaire. Pour la détection de formation tubulaires et de mitoses, les algorithmes de traitement d'images utilisés indépendamment ne sont pas efficaces et l'intervention de l'ontologie est nécessaire pour assurer une segmentation cognitive de telles formations dans l'image. Il y est également proposée une méthode dans laquelle la représentation sémantique décrite dans le langage formel OWL peut être mappée pour représenter l'information dans d'autres langues, afin de relier la sémantique de l'image. Le module de graduation de la région d'intérêt est basé sur l'annotation et la classification pilotée par la représentation sémantique BCGO. Le dernier module fournit le support de validation à l'ensemble du système. Ainsi, en utilisant le BCGO, l'annotation sémantique des images histopathologiques est effectuée, ainsi que l'exploration des lames virtuelles et l'extraction sémantique. BCGO également contribue au processus de graduation en suivant la méthodologie CBIR-CBR proposée dans le chapitre 2, pouvant ainsi fonctionner comme une consultant virtuel pour la graduation.

Chapitre 9. La thèse se termine par une série de conclusions tirées au cours de la synthèse des contributions de l'étude proposée dans le document, et les perspectives de nouvelles recherches dans les ontologies biomédicales, ainsi que l'ouverture vers d'autres caractéristiques de la logique formelle.

En résumé, ces contributions scientifiques sont :

- une étude comparative du point de vue méthodologique (contrairement à d'autres approches qui interviennent au niveau technologique) des deux approches appartenant à différents domaines, le CBIR et CBR, étendues aux services médicaux et qui jette les bases d'une nouvelle méthode hybride

CBIR-CBR pouvant être appliquée dans le processus de graduation du cancer du sein.

- une nouvelle approche rassemblant la représentation sémantique et la représentation de l'image, les deux étant étroitement liés à la perception et la connaissance. Cela a conduit au développement de l'ontologie d''application BCGO (ontologie originale) basée sur une représentation qualitative des concepts de type persistant. La méthodologie de conception d'ontologie (méthodologie en trois étapes) est une méthode générique qui peut être appliquée indépendamment de toute autre représentation. Cette approche présente les avantages suivants: elle résout le fosse sémantique et contextuel - avec une technique d'indexation sémantique à travers des concepts liés a la graduation du cancer du sein, elle assure une forte expressivité et la décidabilité, fournies par deux modules: OWL-DL et les règles SWRL. Enfin, le mécanisme de représentation formelle accompagnée d'un raisonnement sémantique (déduction automatisé normalisé) en BCGO, aide à éliminer le subjectivité sur le BCG. (contribution au niveau de la santé).
- évaluation qualitative de BCGO en utilisant les mesures de granularité, le degré de réflexion de la représentation du domaine, et le degré d'intégration dans d'autres ontologies. Cette évaluation qualitative est accompagnée d'une validation médicale par extraction sémantique, suivie par des résultats de la graduation automatique. L'ontologie est également intégrée dans l'OBO, le but de la réutiliser dans autres ontologies et dans l'ontologie de référence avec laquelle elle est connectée.
- l'intégration de BCGO dans la plateforme microscopie virtuelle, contribuant ainsi au développement de nouvelle orientation - la microscopie virtuelle cognitive MICO par l'indexation sémantique, la capacité de visualisation, l'exploration de lames virtuelles de très grande dimension et l'extraction sémantique.

Les perspectives de recherche ouvertes par la présente étude concernent principalement des questions relatives au raffinement des connaissances*)*, y compris des concepts tels que l'échelle, très important aussi dans la graduation, l'extension de la théorie spatiale avec des relations géométrique et l'expansion de la représentation par des concepts temporels. D'autres méthodes d'évaluation pourraient aussi être explorées, en tenant compte des questions d'ontologies d'application en rapport avec les ontologies de référence, celles-ci pouvant représenter d'autres facteurs importants dans le processus de pronostic. Une autre direction de recherche réside dans la complexité des algorithmes de raisonnement, les questions de implémentation des larges ABoxes et l'analyse de BCGO en utilisant des différents mécanismes d'inférence, tel que celui mis en œuvre dans FACT+ + ou RacerPro.

L'encapsulation des données histopathologique dans un modèle ontologique peut être appliquée à d'autres maladies comme le cancer de la prostate. D'autre part, il peut contribuer à la découverte de nouvelles connaissances en utilisant l'ontologie comme base pour explorer les images. BCGO ontologie peut être intégrée dans un système complexe qui peut contenir plusieurs modèles de systèmes biologiques et d'information au niveau radiologique. Cette direction est exploitée par la communauté VPH (Virtual Physiological Human) qui met l'accent sur le développement et la simulation de modèles biologiques et les représentations

formelles hétérogènes intégrées, ce qui peut contribuer à améliorer les techniques de pronostic, les méthodes de diagnostic et de traitement personnalisé.
Parce que l'essence du Web sémantique est l'ontologie et parce que les ontologies sont en cours d'élaboration dans les domaines médicaux, à la fin de la thèse, il est souligné que la stratégie d'assistance au pronostic basée sur l'ontologie est une direction de recherche de grande importance pour le développement futur des systèmes informatiques comme télépathologie ou le pronostic assiste par ordinateur. Nous nous orientons ainsi vers une génération nouvelle d'outils d'aide pilotes par les ontologies, outils pour lesquels la traçabilité et la reconfigurabilité des systèmes en interaction avec l'utilisateur (en occurrence le médecin ou le pathologiste) prennent une dimension nouvelle.

Résumé en français

Cette thèse aborde l'aide du pronostic basée sur l'image et les ontologies médicales, en utilisant la représentation des connaissances et le raisonnement pour les très grandes images microscopiques. Une application médicale particulière dans laquelle une assistance de type pronostic est nécessaire est la graduation du cancer du sein. Même si cela est considéré comme un outil d'évaluation essentiel dans la pratique de pathologie moderne, les principaux problèmes posés par la procédure manuelle de pronostic sont: la nécessité des connaissances, attention et temps. D'autre part, le manque de représentation sémantique formelle standardisée pour aider l'indexation et la classification de la terminologie, ainsi que l'utilisation d'un mécanisme d'inférence pour assister la graduation représentent des problématiques clé du domaine. Dans ce sens, cette étude propose une représentation formelle qualitative pour la graduation du cancer du sein ainsi qu'une ontologie d'application Breast Cancer Grading Ontology (BCGO) pour décrire les connaissances d'une manière cohérente. Une autre question que nous adressons en proposant l'ontologie, est le fossé sémantique entre les concepts sémantiques de haut niveau et les caractéristiques de l'image de bas niveau. En plus, nous proposons un soutien de théorie spatiale pour la représentation des relations spatiales entre les concepts spécifiques à la graduation du cancer du sein. L'ontologie BCGO est integré dans une plateforme microscopique cognitif virtuelle MICO, pour l'exploration visuelle, l'indexation et l'extraction sémantique de l'image microscopique.

Mots clef: représentation formel, raisonnement qualitatif, Description-Logics formalisme, pronostic base sur l'imagerie médicale, ontologie pour la graduation du cancer du sein, microscopie virtuelle cognitive

Résumé en anglais

This thesis addresses ontology-driven prognosis assistance using knowledge representation and reasoning for very large microscopic medical images. One particular medical application in which prognosis assistance is needed is the breast cancer grading. Although this is considered the key assessment tool in prognosis of modern pathology practice, the main problems of the manual procedure are: time constraints, the need of knowledge and attention. Moreover the lack of formal standardized semantic representation for the indexing and classification of the terminology and the lack of an inference mechanism to assist the grading are key issues of the domain. To this end, we propose a qualitative formal ontological representation of breast cancer grading, an application ontology entitled Breast Cancer Grading Ontology based on OWL-DL and SWRL formalisms. Using this approach, the thesis also tackles the semantic gap between the high- level semantic concepts and the low-level image features. Additionally, we propose a spatial theory support for the representation of the spatial relations between the spatial concepts of the breast cancer grading. This ontology is integrated into a cognitive microscope framework MICO, guiding the image exploration, semantic indexing and retrieval of the microscopic images.

Keywords: formal representation, qualitative reasoning, Description Logics formalism, medical image-based prognosis, breast cancer grading ontology, cognitive virtual microscopy