An Ontology-Based Cancer Diseases Diagnostic Methodology

MARCO ALFONSE, MOSTAFA M. AREF, ABDEL-BADEEH M. SALEM
Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University, Cairo, Egypt
marco_alfonse@cis.asu.edu.eg aref_99@yahoo.com absalem@cis.asu.edu.eg

Abstract: Cancer is a class of diseases characterized by out-of-control cell growth. There are over 200 different types of cancer, and each is classified by the type of cell that is initially affected. This paper discusses the technical aspects of some of the ontology-based medical systems for cancer diseases. It also proposes an ontology-based diagnostic methodology for cancer diseases. This methodology can be applied to help patients, students, and physicians to decide what cancer type the patient has, what is the stage of the cancer and how it can be treated.

Key-Words: Cancer, Ontology, Medical Systems, Computer Aided Diagnosis (CAD), Patient Diagnosis

1 Introduction
An ontology is a formal, explicit specification of a shared conceptualization [1]. This definition identifies four main concepts: an abstract model of a phenomenon termed "conceptualization", a precise mathematical description hints the word "formal", the precision of concepts and their relationships clearly defined are expressed by the term "explicit" and the existence of an agreement between ontology users is hinted by the term "shared" [2]. Ontologies can be used to support a great variety of tasks in diverse research areas such as knowledge representation, natural language processing, information retrieval, databases, knowledge management, online database integration, digital libraries, geographic information systems, visual information retrieval or multi-agent systems [3].

Cancer is one of the most dangerous diseases known to human. The deadliness of this disease can be notified by the fact that cancer caused 13% of all the human deaths in 2007. Cancer, known medically as a malignant neoplasm, is a broad group of various diseases, all involving unregulated cell growth. In cancer, cells divide and grow uncontrollably, forming malignant tumors, and invade nearby parts of the body. The cancer may also spread to more distant parts of the body through the lymphatic system or bloodstream. Not all tumors are cancerous.

Benign tumors do not grow uncontrollably, do not invade neighboring tissues, and do not spread throughout the body. There are over 200 different known cancers that afflict humans [4].

This paper presents an ontology-based cancer diagnostic methodology that can be used to determine the cancer type and stage the patient has and then recommend treatment options. The paper is organized as follows: section 2 gives an background on cancer and ontologies, section 3 presents the related work, section 4 discusses the proposed methodology and finally section 5 contains the discussions and conclusions.

2 Medical Background
Cancer is not a disease; rather it is a group of diseases characterized by uncontrolled growth and spread of abnormal cells. If the spread is not controlled, it can result in death. Figure 1 shows the difference between normal and cancer cells [5]. Cancer is caused by both external factors (tobacco, infectious organisms, chemicals, and radiation) and internal factors (inherited mutations, hormones, immune conditions, and mutations that occur from metabolism). These causal factors may act together or in sequence to initiate or promote the development of cancer. Ten or more years often pass between exposure to external factors and detectable cancer.
Cancer is treated with surgery, radiation, chemotherapy, hormone therapy, biological therapy, and targeted therapy.

Cancer staging describes the extent or spread of cancer at the time of diagnosis. Proper staging is essential in determining the choice of therapy and in assessing prognosis. A cancer’s stage is based on the size or extent of the primary (main) tumor and whether it has spread to other areas of the body. A number of different staging systems are used to classify tumors. A system of summary staging (in situ, local, regional, and distant) is used for descriptive and statistical analysis of tumor registry data. If cancer cells are present only in the layer of cells where they developed and have not spread, the stage is in situ. If cancer cells have penetrated beyond the original layer of tissue, the cancer is invasive and categorized as local, regional, or distant stage based on the extent of spread. Clinicians typically use the TNM cancer staging system, which assesses tumors in three ways: extent of the primary tumor (T), absence or presence of regional lymph node involvement (N), and absence or presence of distant metastases (M). Once the T, N, and M categories are determined, a stage of 0, I, II, III, or IV is assigned, with stage 0 being in situ, stage I being early, and stage IV being the most advanced disease. Some cancers have alternative staging systems (e.g., leukemia). As the molecular properties of cancer have become better understood, tumor biological markers and genetic features have been incorporated into prognostic models, treatment plans, and/or stage for some cancer sites [5].

3 Related Work
Ontologies can be used to diagnose cancer diseases. To build an ontology is a similar process than to build an object oriented program, however, classes and objects in a program are about data structures, whereas classes and objects in ontologies are about the domain. A good practice to support ontology building process is to produce an ontology specification document written in natural language with information such as: the purpose of the ontology, its end users, use case scenarios, degree of formality used to codify the ontology and its scope. An ontology specification document should be conciseness, (it must include just relevant terms without duplicates), partial completeness and realism, (meanings of the terms and their relationships making sense in the domain) [7]. Figure 2 shows a general methodology to build an ontology.

Many methodologies, languages and tools are used to support ontological engineering process. In [8] the authors provide the different methodologies used to build ontologies, the languages used to codify ontologies and finally the tools that can be used to support ontology building process.

In [9] the authors provide the computerization and execution of a breast cancer follow-up clinical practice guideline. The computerization of the clinical practice guideline led to the development of
a breast cancer ontology. They present their breast cancer ontology which models the knowledge inherent within the breast cancer follow-up clinical practice guideline - the breast cancer ontology serves as the knowledge source to determine patient-specific recommendations. They discuss the ontology engineering process that highlights the specification of their breast cancer ontology in terms of clinical concepts and the relationships between the concepts expressed as OWL classes and properties, using the protege ontology development tool.

In [10] the authors introduce a novel Ontology-based Argumentation Framework (OAF) that links a logic-based argumentation formalism and description logic ontologies. They show how these two formalisms can be tightly coupled by observing a few simple restrictions, and provides features not available in either formalism alone. Their work is evaluated in a large case study on decision-making in treatment choice in breast cancer, where rules are developed from the results of published clinical trials, and they present a small subset of this to demonstrate the use of the system.

In [11] the authors present an Ontology-based Modeling of Clinical Practice Guidelines: A Clinical Decision Support System for Breast Cancer Follow-up Interventions at Primary Care Settings

The ASSIST project aims to facilitate cervical cancer research by integrating medical records containing both phenotypic and genotypic data, and residing in different medical centres or hospitals. The goal of ASSIST is to enable the evaluation of medical hypotheses and the conduct of association studies in an intuitive manner, thereby allowing medical researchers to identify risk factors that can then be used at the point of care to identify women who are at high risk of developing cervical cancer. In [12] the authors present the current status of the ASSIST medical knowledgebase. In particular, they discussed the challenges faced in constructing the ASSIST integrated resource and in enabling query processing through a domain ontology, and the solutions provided using the AutoMed heterogeneous data integration system. They focus on data cleansing issues, on data integration issues related to integrating relational medical data sources into an independent domain ontology and also on query processing. Of particular interest is the challenge of providing an easily maintainable integrated resource in a setting where the data sources and the domain ontology are developed independently and are therefore both highly likely to evolve over time.

In [13] a breast cancer ontology has been built and gathers the terms used by lay people to talk about breast cancer. The resulting ontology has been the core of a health consumer query reformulation application. This work proposes the use of spreading activation technique through the ontology to infer new concepts from the ones initially identified in a health consumer question.

In [14] the authors present a benign/malignant breast cancer classification model based on a combination of ontology and case-based reasoning to effectively classify breast cancer tumors as either malignant or benign. This classification system makes use of clinical data. Two CBR object-oriented frameworks based on ontology are used jCOLIBRI and myCBR.

Table 1 shows the different ontology based medical systems for cancer diseases currently exist.

<table>
<thead>
<tr>
<th>System</th>
<th>Ontology Used</th>
<th>Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Cancer Follow-Up Clinical Practice Guideline</td>
<td>Breast Cancer</td>
<td>Clinical Decision Support System</td>
</tr>
<tr>
<td>Argument-Based Decision-Making in Breast Cancer</td>
<td>Breast Cancer</td>
<td>Links a logic-based argumentation formalism and description logic ontologies</td>
</tr>
<tr>
<td>A Clinical Decision Support System for Breast Cancer Follow-up</td>
<td>Breast Cancer</td>
<td>Decision Support System</td>
</tr>
<tr>
<td>The ASSIST project</td>
<td>Cervical Cancer</td>
<td>Data integration approaches</td>
</tr>
<tr>
<td>Ontology-Based Reformulation of Health Consumer Queries</td>
<td>Breast Cancer</td>
<td>Spreading Activation Techniques through the ontology</td>
</tr>
<tr>
<td>A Breast Cancer Classifier</td>
<td>Breast Cancer</td>
<td>Combination of ontology and case-based reasoning</td>
</tr>
</tbody>
</table>

Table 1: The Ontology-Based Medical Systems for Cancer Diseases
4 The Proposed Methodology

The proposed methodology contains three basic modules namely; the diagnostic module, the staging module and the treatment recommendation module. In order to detect patient disease, the patient provides his/her signs and symptoms to the diagnostic module, which detects what type of cancer the patient has. Once the type of the cancer is determined, the staging module finds the current stage of the cancer based on the cancer type and the signs and symptoms provided by the patient. Based on the determined cancer type and cancer stage, the treatment recommendation module can recommend a specific treatment for the case at hand.

All the three modules interact with a database of cancer ontologies through the query module, which maps from the query of the asking module to the structure of the vocabulary of the ontologies stored in the database of cancer ontologies. The database of cancer ontologies describes the different types of cancer diseases in detail. Each cancer ontology describes the cancer in terms of its structure, signs and symptoms, staging and treatment. Figure 3 shows the proposed methodology.

Currently the authors already developed three medical ontologies describing three different cancer diseases, which are the lung cancer ontology [15], the breast cancer ontology [16], and the liver cancer ontology [17]. These ontologies was developed using the protégé-OWL tool and encoded in OWL-DL format.

Each module (diagnostic, staging or treatment recommendation) can be applied as a stand-alone application according to the user requirements. For example, if the user knows the cancer type and wants to determine the cancer stage it can use the staging module by providing the cancer type and signs and symptoms and the module will determine the current stage and so on.

This methodology can be evaluated in terms of its modules. The application of this methodology will allow evaluating the diagnostic module, the staging module and the treatment module. Expert physicians and medical datasets will achieve the evaluation of this methodology.

This proposed methodology can be applied to provide patients, students of medicine and physicians to what is called second opinion to decide what cancer type the patient has, what is the stage of the cancer and how it can be treated.

5 Conclusion and Future Work

This paper discusses the technical aspects of the existing ontology-based medical systems for cancer diseases. Most of these systems used for diagnosis of the cancer disease. These systems use different approaches for the diagnosis of the cancer disease (e.g. decision support system, data integration, ontology …etc). The paper also proposes a new ontology-based cancer diseases diagnostic methodology. This methodology can be applied for the diagnosis of cancer type, determining the stage of the specific cancer and it can recommend treatment solutions. It can be used by patients, students and physicians to decide the type of the cancer the patient has, the stage of the cancer and the treatment options. The main contribution of this methodology is its capability to provide the treatment recommendation for the case at hand. Currently the database of cancer ontologies has only three types of cancers, in future, it is planned to cover all types of cancer ontologies.

Figure 3: The Proposed Methodology
References: