

# Ontology Based Information Retrieval of Cancer Diseases

**Cho Thet Lwin**

Computer University (Taung Ngu)

[t.mienge@gmail.com](mailto:t.mienge@gmail.com), [chothetlwin.ctl@gmail.com](mailto:chothetlwin.ctl@gmail.com), [cutaungngu2009@gmail.com](mailto:cutaungngu2009@gmail.com)

## Abstract

*The proposed system has being specifically focuses on the development of a cancer information system by using information retrieval (IR) techniques. Information Retrieval System should be able to identify the intent of information search as well as the content of information source in order to retrieve the items that can satisfy the information need. Ontology has become common on the World-Wide-Web. Ontology provides a useful technology for semantic information. It contains the seeds of important applications, but to realize its full potential it must be linked into a single global system. The OWL language is a major step towards developing that potential. The system is implementing by using Java programming language.*

## 1. Introduction

People from all over the world use the internet as the communication media for accessing information they need and giving information they want to deliver. Ontology has become common on the World-Wide-Web.

Information Retrieval (IR) deals with the representation, storage, organization of and access to information items. The representation and organization of the information items should provide the user with easy access to the information in which he is interested.

In this thesis, user can search the cancer disease symptoms and other related disease information. Cancer is a disease that needs to be treated seriously of community-centric cancer information system and one of the inadequacies of HTML lies in its ability to support fast information searching. This project tries to solve the problems stated above, that is, to develop Ontology-based cancer information system that will help user, especially, to search for cancer information. Besides that, IR techniques will be used to enhance the system with full-text searching capability in order to increase the effectiveness of information searching. Keywords used during the searching are highlighted in the retrieved documents in order to help user to locate them quickly.

The system to develop ontology based information retrieval of cancer disease. OWL language is used in this system. A large number of organizations have been exploring the use of OWL, with the many tools currently available. OWL is a W3C recommendation language. OWL (Web Ontology Language) adds more classes, cardinality, and equality, richer typing of properties and enumerated classes.

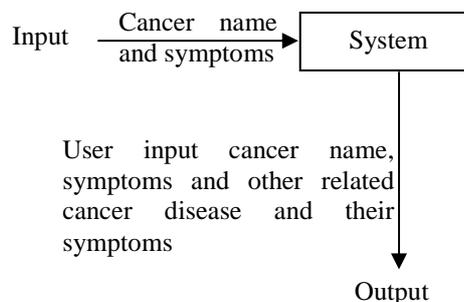
## 2. Information Retrieval System

Information retrieval is the science and art of locating and obtaining based on information needs expressed to a system in a query language. In general, the search process is accomplished by specifying one or more keyword to search for (associated with, e.g., Boolean logic), which the content or metadata.

An information retrieval system is to be distinguished from a data retrieval system. Information Retrieval usually deals with natural language text which is not always well structured and could be semantically ambiguous. On the other hand, a data retrieval system deals with data that has well defined structure and semantics.

An information retrieval system attempts to find relevant documents to respond to a user's request. An information retrieval process begins when a user enters a query into the system. Queries are formal statements of information needs.

User inserts the input data such as cancer name and symptoms. Output is the related other cancer disease and their symptoms.



**Figure 1: Typical Information Retrieval System**

### 3. Ontology

Ontology is an important emerging discipline that has the huge potential to improve information organization, management and understanding. Ontology is the term referring to the shared understanding of some domains of interest, which is often conceived as a set of classes (concepts), relations, functions, axioms and instances. Ontologies have been used in database and information retrieval areas as a support for distributed and heterogeneous data sources interoperability. In a database point of view, ontology may be considered as a partial domain specification, describing basically concepts and relationships between them. When data is marked up using ontologies, software agents can better understand the semantics and therefore more intelligently locate and integrate data for a wide variety of tasks.

Ontology are used by people, databases, and applications that need to share domain information (a domain is just a specific subject area of area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc.).

Ontology is a complex multi-disciplinary field that draws upon the knowledge of information organization, natural language processing, information extraction, artificial intelligence, knowledge representation and acquisition. Ontology as the shared and common understanding of the domain that can be communicated between people and application systems, has a significant impact on areas dealing with vast amounts of distributed and heterogeneous computer-based information, such as World Wide Web and internet information systems, complex industrial software applications knowledge management, electronic commerce and e-business.

#### 3.1. Web Ontology Language

The Web Ontology Language (OWL) is a semantic markup language for publishing and sharing ontologies on the World Wide Web. OWL is developed as a vocabulary extension of RDF and is derived from the DAML+OIL Web Ontology Language.

The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schemas (RDFS) by providing additional vocabulary along with a formal semantics, OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL and OWL Full.

Information in OWL is gathered into ontologies, which can then be stored as documents in the World Wide Web. One aspect of OWL, the importing of ontologies, depends on this ability to store OWL ontologies in the Web.

#### 3.1.1. The term use for OWL

- Concepts are known as classes in owl and are created with the owl:Class.
- A class may contain its documentation (with rdfs:comment).
- rdfs:subClassOf contains class expressions. It allows defining the superclasses of the class.
- owl:disjointWith that the class cannot share with the other class.
- Equivalence between classes can be defined with the OWL: equivalentClass.
- owl:oneOf defines a class by enumerating all its instances.
- owl:intersectionOf, owl:unionOf and owl:complementOf define a class expression as a conjunction, a disjunction, or a negation of other class expressions respectively.
- OWL documents are usually called OWL ontologies, and are RDF documents.
- the root element of a OWL ontology is an rdf:RDF element which also specifies a number of a namespaces.

A typical OWL ontology begins with a namespace.

#### 3.1.2. The Cancer and Symptoms are used in this System

The system will solve six kinds of cancer diseases. They are Blood Cancer, Breast Cancer, Lung Cancer, Liver Cancer, Kidney Cancer and Prostate Cancer.

The symptoms contain in the blood cancer are fatigue, malaise, breathlessness, weakness, excessive or easy bruising, bleeding gums or frequent nose bleeds, recurrent infections or fever, sweating at night, weight loss, anorexia, lymph node (gland) enlargement, lumps or abdominal distension due to enlarged abdominal pain organs and abdominal pain. And then back pain, bone pain, bone fractures from minimal trauma, confusion, delirium, headaches, visual disturbance and decreased urination.

The symptoms for breast cancer are a lump or thickening in an area of the breast, a change in the size or shape of a breast, dimpling of the skin, a change in the shape of your nipple, particularly if it turns in, skins into the breast or becomes irregular in

shape, a blood-stained discharge from the nipple, a rash on a nipple or surrounding area and a swelling or lump in your armpit.

The lung cancer contain aching chest, shortness of breath, weight loss, finger clubbing (broadening or thickening of the tips of the fingers), hoarse voice, swollen glands (lymph nodes), chest and pain.

The liver cancer has the following symptoms. They are unintentional weight loss, loss of appetite, pain and/or discomfort on the right side of the abdomen, pain or discomfort that occurs in the right shoulder blade area.

The kidney cancer has occurs these symptoms. They are chronic fatigue, rapid weight loss, leg and ankle swelling, hypertension (high blood pressure), fever, presence of blood in urine (seen either by the eye or microscopically) and pain in side or lower back.

These symptoms can occur the prostate cancer. They are burning or pain during urination, weight loss, inability to urinate, frequent nocturnal urination, weak urine stream, blood in urine (either seen by the eye or microscopically), pelvic pain, back or hip pain.

### 3.1.3. Common symptoms of Cancer

Some cancer symptoms are:

- Persistent fatigue: Fatigue is one of the most commonly experienced cancer symptoms.
- Unintentional weight loss: While it may be a welcome surprise to lose weight without trying, it can be a red flag for many illnesses, including cancer.
- Pain: Typically, pain is cancer symptom of cancer, except in some cancer types like those that spread to the bone. Pain generally occurs when cancer spreads and begins to affect other organs and nerves.
- Fever: In relation to cancer, a fever that is persistent or one that comes and goes frequently can signal stress on the immune system. Fevers are commonly associated with types of cancer that affects the blood, like leukemia and lymphoma, but are also common in people whose cancer has spread.
- Bowel changes: These symptoms are most commonly associated with colon cancer, but are also related to other cancer types.
- Chronic cough: Blood and/or mucus may accompany the cough and can be caused many conditions. In relation to cancer, a chronic cough with blood or mucus can be symptom of lung cancer.

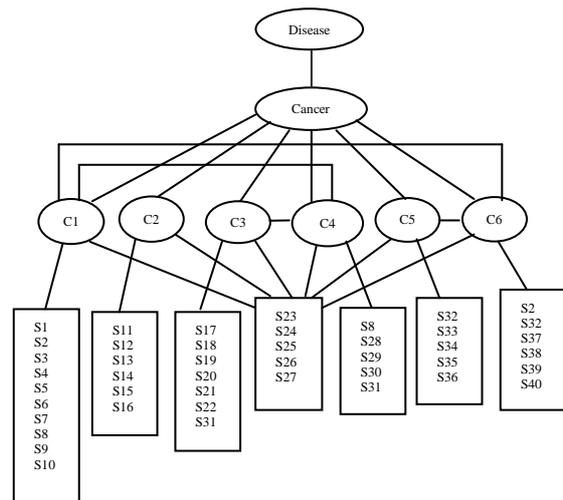


Figure 2: Concept Hierarchy for Cancer Diseases

symbols	cancer name
C1	Blood cancer
C2	Breast cancer
C3	Lung cancer
C4	Liver cancer
C5	Kidney cancer
C6	Prostate cancer

Table 1: Table for Cancer Symbols of System

symbols	symptoms name
S1	abdominal pain
S2	back pain
S3	bone pain
S4	confusion
S5	decreased urination
S6	excessive or easy bruising
S7	headache
S8	malaise
S9	weakness or lack of energy
S10	frequent nose bleeds
S11	blood-stained discharge from the nipple
S12	change in breast size or shape
S13	rough or scaly patches
S14	rapid or irregular heartbeats
S15	rash on nipple or surrounding area
S16	swelling in armpit
S17	figure clubbing
S18	hoarse voice
S19	shortness of breath
S20	shoulder glands
S21	shoulder pain
S22	stool contain blood
S23	cough
S24	fever
S25	fatigue
S26	weight loss
S27	pain
S28	decreased appetite

S29	pain on abdomen right
S30	sore throat
S31	aching chest
S32	blood in urine
S31	chronic fatigue
S34	high blood pressure
S35	leg and ankle swelling
S36	pain in side or lower back
S37	frequent urination
S38	burning or pain during urination
S39	weak urine stream
S40	hip pain

**Table 2: Symptoms of the use of System**

symbols	common symptoms
S23	cough
S24	fever
S25	fatigue
S26	weight loss
S27	pain

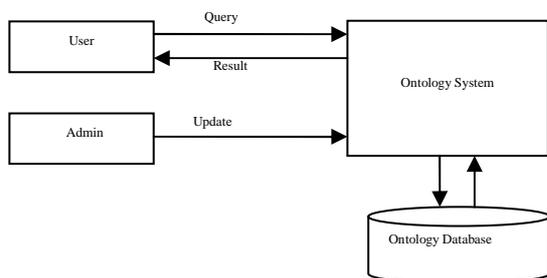
**Table 3: Common symptoms for use of system**

#### 4. Overview of the System

This system is implemented to search concept and creating about the concept. It develops ontology concept by using Ontology Web Language (OWL) schema repository.

In this system contain two portion, they are user and admin. User can query cancer information to the ontology system and the system reply to user's need query information result.

Admin can update the symptoms and new cancer disease name to the ontology system. The system can find the concept in the ontology database.



**Figure 3: Overview of the System**

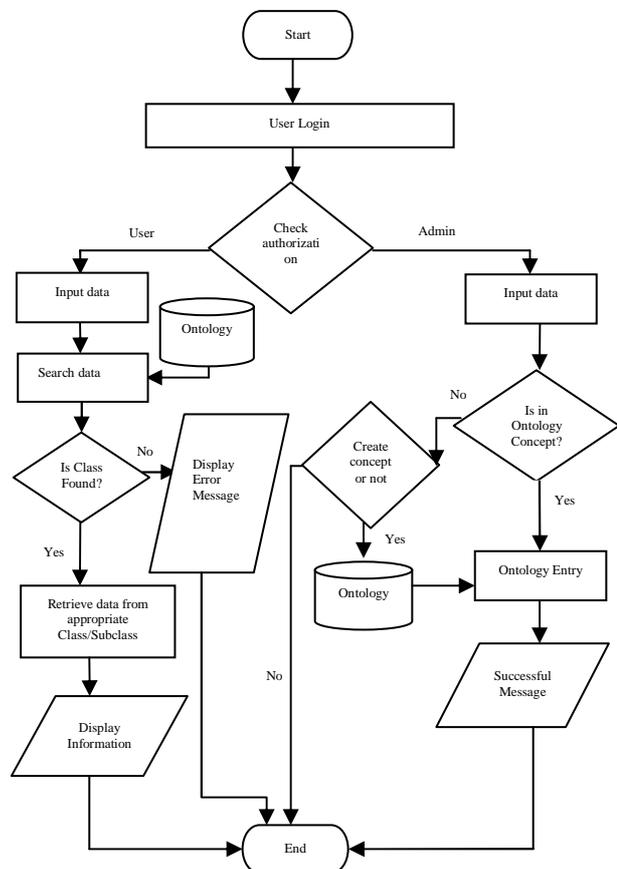
#### 4.1. The propose system architecture

The data flow diagram of the system is in figure 4. Firstly, the system starts the user login. And then the system check the authorization which the

user is admin or not. If the user is admin, user inserts the input data such as cancer name and symptoms. Then system checks the input data is ontology concept or not. Input data is ontology concept, user can entry in the ontology database. And show successful message and end the system.

If the input data is not the ontology concept and the system checks the user is create ontology concept or not. The user wants to create ontology concept in the ontology database and entry the ontology. The user wants not the create concept the system is finish.

The user is not admin, the user insert the input data and search the data in the ontology database. The class is found, user can retrieve the data from appropriate class and subclass and display the relevant information and end the system. If the class not found the system display error message and end the system.



**Figure 4: System Flow Diagram**

#### 4.2. System Implementation

User inserts the input data such as cancer name and symptoms. In this system, all symptoms are concept relations between the various cancer diseases' symptoms. For example, user inserts the

cancer name is blood cancer and type the symptoms for the blood cancer. And then the system checks the symptoms because these symptoms are include other cancer disease or not.

If the symptoms are include other cancer diseases, at that time, the system display the other cancer name and related their symptoms. This system is intend to get the user about the cancer's symptoms and other related diseases' in this system.

```
<owl:Class rdf:ID="blood">
  <rdfs:subClassOf rdf:resource="#cancer"/>
</owl:Class>
<owl:Class rdf:ID="symptoms">
  <rdfs:subClassOf
rdf:resource="symptomsDescription"/>
  <owl:oneOf rdf:parseType="Collection">
    <owl:Thing
rdf:about="#fatigue"/>
    <owl:Thing
rdf:about="#weakness"/>
    <owl:Thing rdf:about="#bleeding
gums or frequent nose bleeds"/>
    <owl:Thing rdf:about="#sweating
at night"/>
    <owl:Thing rdf:about="#bone
pain"/>
    <owl:Thing
rdf:about="#headaches"/>
  </owl:oneOf>
</owl:Class>
```

## 5. Conclusions

The proposed of this system is very useful in concept creation with easy and quick way. In addition, it can search cancer symptoms easily. Users can exactly access the necessary information for input data. Ontology is the most powerful way to determine concept for the categorized system.

The system gathering symptoms for developing ontology based information retrieval is done manually. It is really misunderstanding for novice users. They need to understand OWL and how to create Ontology concept. And then, this system is developed for only six cancer disease symptoms and related disease and their symptoms. So the user wants to insert new other cancer and their symptoms. And then other resources are wanted to process, the system will be modified.

## 6. References

[1] Banerjee, Jay, Golbeck, Jennifer, Hendler, James and Kalyanpur, Aditya, "OWL: Capturing Semantic Information Using a Standardized Web Ontology Language", University of Maryland.

- [2] Davies, John; Fensel, Dieter; Harmelen, Frank van "TOWARDS SEMANTIC WEB: Ontology-driven Knowledge Management" ISBN 0-470-84867-7, John Wiley & Sons Ltd, England (2002).
- [3] Ding, Ying; Foo, Schubert. "Ontology Research and Development, Part 1-A Review of Ontology Generation". Journal of Information Science, Nanyang Technological University, Singapore (2002).
- [4] Ernest J.Friedman-Hill, Sandia National Laboratories "JESS, the Rule Engine for the Java Platform", Version 7.0p1 (21 December 2006) DRAFT.
- [5] Guarino, N., Carrara, M. and Giaretta,P.1994.. Formalizing Ontological Commitment. In Proceeding of National Conference on AI (AAAI-94). Seattle, Morgan Kaufmann.
- [6] Ikeda, Mitsuru; Mizoguchi, "Toward ontology engineering" Technical Report, Osaka University.
- [7] Noy, Natalya F. and McGuinness, Deborah [L.noy@smi.stanford.edu](mailto:L.noy@smi.stanford.edu) and [djm@ksl.stanford.edu](mailto:djm@ksl.stanford.edu) "A Guide to Creating Your First Ontology" Stanford University, Stanford, CA, 94309
- [8] OWL Web Ontology Language Reference, W3C Recommendation 10 February 2004 <http://www.w3.org/TR/2004/REC-owl-ref-20040210/>.