2009 Ninth International Conference on Intelligent Systems Design and Applications

Designing a Conceptual Model for Herbal Research Domain using Ontology Technique

Azlida Mamat
Department of Information System
Faculty of Computer Science & Information Systems
Universiti Teknologi Malaysia, Skudai, Johor
azlida@utm.my

Azizah Abdul Rahman
Department of Information System
Faculty of Computer Science & Information Systems
Universiti Teknologi Malaysia, Skudai, Johor
azizahar@utm.my

Abstract—**This paper demonstrates a conceptual model for herbal research domain using ontology technique. Important information in herbal research domain have been identified through interview sessions conducted on groups of herbs researchers in fields of engineering, medical and botany. The variation of information gathered led to the difficulty of database design. However, we propose a methodology for the development of conceptual model using ontology technique, with adoption of four extended relationship components. It covers relationships of temporal, prerequisite, mutually inclusive and mutually exclusive. In order to ensure the conceptual model is well defined, a set of rules for keyword searching is created to verify preciseness of output produced. The rules created in this paper will be executed on Herbal Research E-Centre prototype.**

Keywords-component;; Ontology; Conceptual Model; Database Design

I. INTRODUCTION

All Malaysian government is using the five-year Ninth Malaysian Plan to exploit the full potential of biotechnology. Malaysia wants to gain and capitalise on its biodiversity through commercialization of discoveries in health related natural resources or natural products. There are four categories of natural products which are plant, microbial, animal and marine. This paper focuses on plant, specifically on herbs. People awareness toward valuable herbs is evolving over the time because as the usage of herbal products has increased rapidly. This is due to introduction of many herbal products in the market and the popularity of herbal usage. Despite a lot herbal products launched in market, herb consumers are still feel doubtful to of its application asy herbal medicine. They are worried about the effectiveness and side effects of the products. Based on this scenario, many parties in herb industry have started conducting research on herbs. They study and explore the plant part values that are beneficial to the human beings. Some of them work hard to extract and analyze the healing components from herbal plants, to get as much profit as possible from herbal products without thinking about the knowledge from herbal research that should be shared to others[1]. Nowadays, too many launched herbal portals are still ineffective in helping users to obtain the herbal research information. This is due to unstructured and scattered information. Therefore, we take an initiative to develop a Herbal Research E-Centre (HREC) as a centralized herbal research information that can be used by Malaysian herb community. HREC plays the role as a repository of herbal research which is allows herbs researchers to store and retrieve information of on herbal research. Research network among herbs researchers can be developed through this system. However, the effectiveness of this system can be achieved with the cooperation of herbs researchers to contribute their research findings into this system. HREC is expected to become complete and stable over the time through ontology verification by domain expert.

The community in herb industry consists of entrepreneurs, suppliers, special services, manufacturers, researchers, and etc. This paper focuses on researcher community. There are several fields in herbal research domain which are medical, engineering and botany. In order to identify information that are required by herb researchers, and the ones is information that herbs researchers are willing to share, interview sessions have been conducted on focus groups from the three mentioned fields. The information has been organized and analyzed using ontology technique. Based upon the ontology created, author tries to extend role of ontology in designing database. This paper demonstrates how ontology representation can assist database design. Several related research of component on database design described in Section 2. Section 3 presents methodology that has been used in designing conceptual model. Section 4 shows results based on steps in methodology. Section 5 demonstrates rules for keyword searching to test capability of relationship components in representing herbal research domain.

II. RELATED RESEARCH

A. Database Design

Database design is a process to produce detailed data model of a database. The detailed data model consists of detailed value parameters, attributes, primary key, foreign key and relationship between entities. The designing of the database needs an excellent developer's understanding of

978-0-7695-3872-3/09 $26.00 © 2009 Crown Copyright
DOI 10.1109/ISDA.2009.192

both two criteria which are the domain area and database development. Effective database design can assist developer to perform well from the beginning. In addition, it can reduce costs and time during development process. An excellent database development is important to get an optimal performance and highly productivity[2]. In order to achieve the quality of system, the designer should work hard in representing information through database design to ensure the database can works properly.

B. *Conceptual Modeling*

Conceptual modeling is a process to model data of domain. Conceptual modeling is a well known technique of data modeling. It represents domain entities, meaning of the data, concepts or terms used by domain experts, function or relationship between concepts. Conceptual model or also known as conceptual level schema is a part of the process in database design which determinesd information needs of user[3] . It is able to provide an accurate, complete representation of someone's understanding of the domain, with adaptation for different purposes[4].

C. *Ontology*

Ontology has been developed to improve the level of information organisation, management and understanding[5]. Ontology has been established for knowledge sharing and is widely used as a means for conceptualizing structuring domains of interest[6]. In computer and information sciences context, ontology is a set of representational primitives with which to model a domain of knowledge. The representational primitives are typically refers to classes (or sets), attributes (or properties), and relationships (or relations among class members)[7]. The concepts and idea in ontology is designed based on developer's understanding of problem domain, and their creative process. There is no single correct ontology for any domain and no two ontologies designed by different people are would be the same[8]. In the herbal research domain, we apply the ontology technique to organize variety and unstructured information gathered from fields of engineering, medical and botany.

1) *Herbal Research Ontology*

During ontology development, there are three key challenges that we need to tackle. The key challenges; 1) overlapping content, 2) variety term used, and 3) complexity relation among term[9]. However, one by one key challenges have been solved through ontology based approach. The content overlapping is tackled through identification class and subclass. The problem of variety term used is settled through identification synonym term. Last key challenge is complexity of relation is simplified through assigning four relationship component provided by Sugumaran and Storey[10].

III. METHODOLOGY FOR DATABASE DESIGN USING ONTOLOGY BASED APPROACH

This section explains the methodology in for the designing a conceptual model using ontology representative components. The proposed methodology contains consists of 9 steps, with adoption of four component relationships provided by Sugumaran and Storey[10].

Common ontology representation or basic relationships for conceptual modeling are *is_a*, *synonym*, and *related_to*[10]. This research applies four extended relationship component in order to solve an inability of basic relationship in modeling domain knowledge. The purpose of this application is to simplify in defining the rules exist in herbal industry. The following four types of relationship component are *Prerequisite, Temporal, Mutually Inclusive and Mutually Exclusive.*

Prerequisite relationship concerns on one term that depends upon another. For example in the herbal research, a testing requires an herb. However, information about herb can exist without requiring testing. Let say there have two terms; A and B. If A requires B: A → B, it means A prerequisite to B.

Temporal relationship refers to scenario that one term must occur before another. Let say there have two terms; A and B. If A must precede B, the temporal constraints is A |→ B, However, temporal relationship quite hard to be establish because it concerns on dimensions of time.

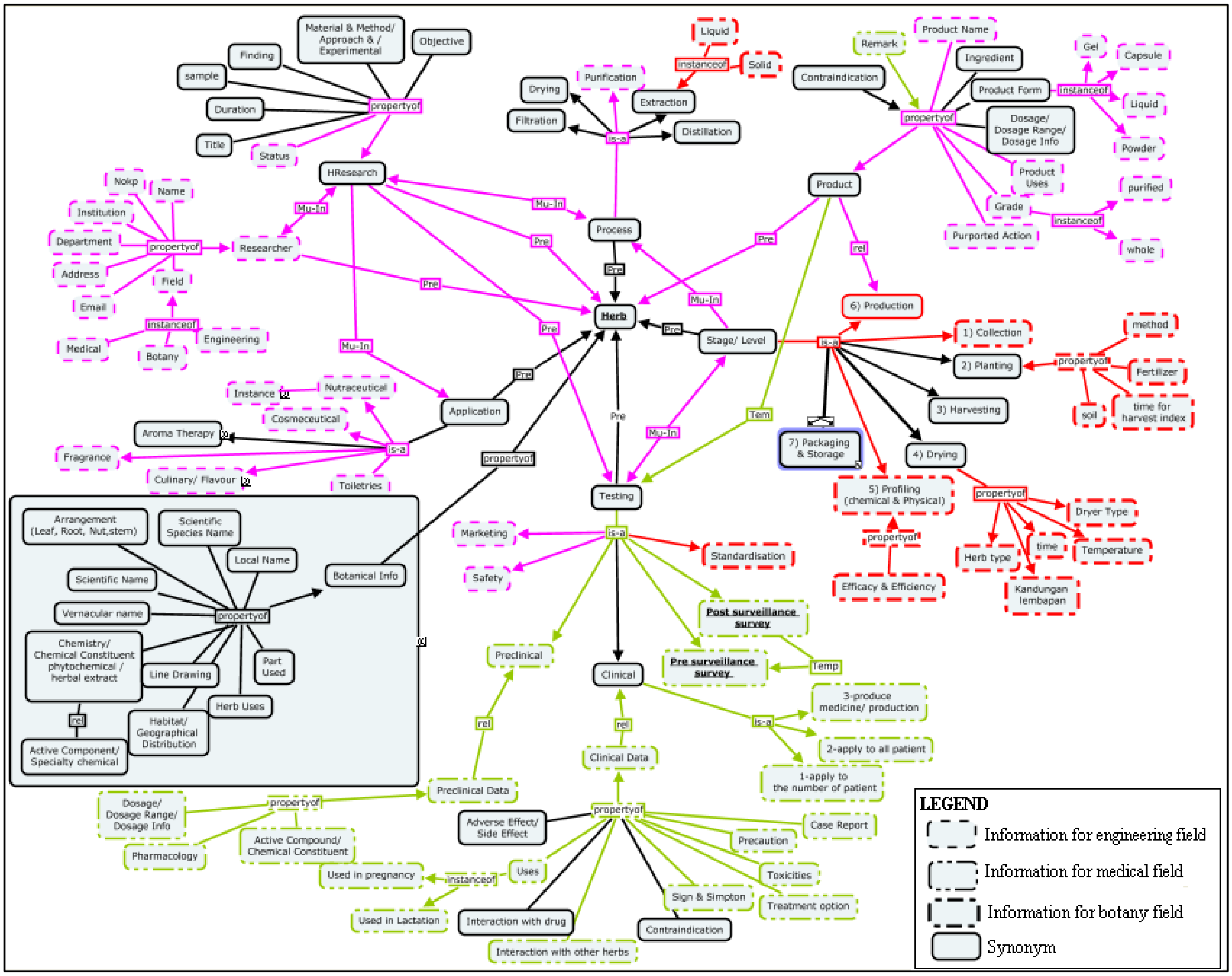
Mutually Inclusive relationship focuses on one term that requires another term for its existence. Let say there have two terms; A and B. If A requires B and B also requires A; A B. For example in the herbal research domain, a process requires an application and an application can not be produced without a process.

Mutually Exclusive relationship. One term cannot occur at the same time. Let say there have two terms; A and B. The A and B cannot operates in the same time; A B. In the herbal research context, an individual cannot be the administrator and contributor of the same time. This is to avoid bias during validation process.

There are nine steps to model the conceptual of this domain. The initial four steps are common steps for developing an ontology provided by Natalya and Deborah[8]. Step five involves adaptation of existing relationship into four new extended relationship proposed by Vijayan Sugumaran and Veda C.Storey. They highlight that "incorporation and use of ontologies is effective in creating entity-relationship models. They produces a prototype called Ontology Management and Database Design Environment which supports entity relationship modeling activities[10]. In order to know the capability of the components, we need to adapt these components into herbal research domain.

The whole steps are:

a) *Enumerate all term in the domain*
b) *Identify term (definition, synonym word, related, hypernym)*
c) *Define classes, subclass and class hierarchy.*
d) *Identify the relationship between classes.*
e) *Adapt the existing relationship to four extended relationship component[10].*
f) *List all classes of each component*



g) Create the primary key for each class.

Figure 1. Herbal research ontology in engineering, botany and medical fields

h) Create the foreign key for classes based on component of relationship.

i) Identify the attribute or properties for each class.

IV. RESULT

This section describes results based on steps for conceptual modeling as given in Section 3. The herbal research ontology created based on Step1 to 5 is visualized in Figure 1. The herbal research ontology is developed using the Cmap tool. The herbal research ontology covers information from three fields of engineering, medical and botany. Four types of patterns are represented. It represents the information from fields of engineering, medical, and botany and synonym term (refer legend).

The Step 6 is a listing of all classes of each component. Based on figure 1, there are 8 classes and 25 subclasses have produced based on author understanding of the domain. However, not all classes in Figure 1 presented in this paper. In this paper, we present two classes of each component.

Table 1 shows classes for each component. Prerequisite component shows prerequisite relationship between classes of *Testing*, *Level* and *Herb*. It means class *Testing* and class *Level* is depending on class *Herb,* but the existence of class *Herb* are not depends on class *Testing or* class *Level*. Mutually inclusive means objects that mutual depends on another. For this domain, class *HResearch* depends on class *Researcher*, and class *Researcher* depend on class *HResearch*. Temporal component describes objects that dependence based on dimension of time. In this domain, Class *Postsurveillance* cannot proceed before class *presurveillance* has finished. Basic component *is_a* and *related_to* exist in majority of domain.

Table 2 shows primary key created for each class. A primary key is used to uniquely identify each row in a table. In coding process, the primary key is playing a role as key reference which can identify the information needed by input user. A primary key cannot be null and the value assigned to a primary key should not change over time.

TABLE I. LIST OF CLASSES

Prerequisite ***Class Testing→ Class Herb*** ***Class Level → Class Herb***
Mutually Inclusive ***Class hresearch→ Class Researcher*** ***Class Level → Class Process***
Temporal ***Class Post-surveillance→ Class Pre-surveillance***
Is_a ***Class Process*** ***- Subclass Distillation*** ***- Subclass Extraction***
Related_to ***Class Product → Class Production***

This section describes the details on how primary key of related classes can be a foreign key for others related class. Creating foreign key of classes based on component of relationship.

TABLE II. LIST OF CLASSES WITH PRIMARY KEY

Classes	Primary Key
Herb	***idherb***
Testing	***idtesting***
Level	***idlevel***
Herbal Research	***idHR***
Researcher	***idresearcher***
Process	***idprocess***
Postsurveillance	***idpostsurvey***
Presurveillance	***idpresurvey***
Distillation	***iddistill***
Extraction	***idextract***
Product	***idproduct***
Production	***idpro***

A way to create a foreign key for classes is different based on the types of relationship components. In order to simplify results for this step, we describe the correlation between classes from Table 1 through creation of foreign key. In this case, there may have more than one foreign key in a class. This existence of the foreign key is based on relationships they have.

TABLE III. FOREIGN KEY FOR EACH CLASS

Component : Mutually Inclusive ***Class Level*** — ***PK= idlevel***, ***FK= idtesting*** ***Class Testing*** — ***PK= idtesting***, ***FK= idlevel*** ***Class Herbal Research*** — ***PK= idHR***, ***FK= idresearcher*** ***Class Researcher*** — ***PK= idreseracher***, ***FK= idHR***	**Relationship Component : Pre-requisite** ***Class Testing*** — ***PK= idtesting***, ***FK= idherb*** ***Class Herb*** — ***PK= idherb*** ***Class Level*** — ***PK= idlevel***, ***FK= idherb*** ***Class Herb*** — ***PK= idherb***
Relationship Component : Temporal ***Class Pre-Surveillance*** — ***PK= idpretsurvey*** ***Class Post-Surveillance*** — ***PK= idpostsurvey***, ***FK= idpresurvey***	**Relationship Component : Related To** ***Class Product*** — ***PK= idproduct***, ***FK = idpro*** ***Class Production*** — ***PK= idpro***, ***FK= idproduct***

Table 3 shows result from step 8. *PK* represent as Primary key while *FK* represent as Foreign key. Next section is explanation on how to assign foreign key based on relationship components.

First, the prerequisite component relationship is referring to the object that depends on another object. For example class *Testing* prerequisite class *Herb*. It means testing information requires herb information. In herbal research domain, majority of herb researcher find testing information based on herb. Let say, we want to know about preclinical testing of Eurycoma Longifolia or well known as Tongkat Ali. In this scenario, testing information is depending on herb information. In order to produce correct output, primary key of class *Herb* must be exist in class *Testing*. This is due to simplify information retrieval of class *Testing*.

Second, mutually exclusive relationship represents objects that mutually dependence. For example, herbal research cannot exist without researcher, and people cannot be a researcher without herbal research. It means the existence of two mentioned classes is depends upon another. In this scenario, we assign foreign key for each related class is a primary key of its pair. It means foreign key of class Researcher is idHR, and foreign key for class *HResearch* is idresearcher.

Third, temporal relationship refers to the existence of objects in not in the same time. In creation creating a foreign key, we assign foreign key of class *Postsurveillance* is idpresurvey, and there is no foreign key of class *Presurveillance*. This is due to postsurveillance testing cannot be executed until presurveillance completed.

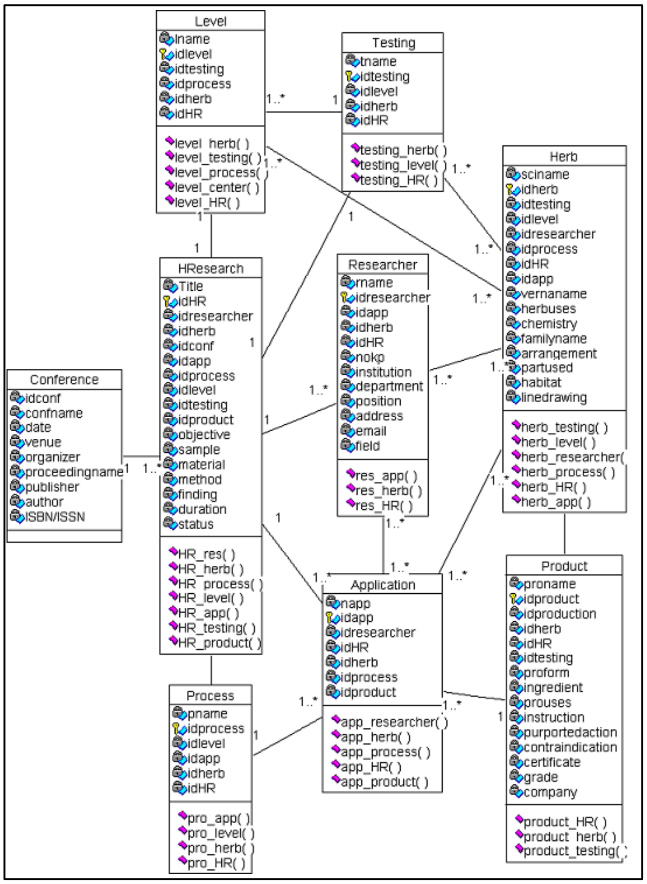
Mutually exclusive are not applicable or can not be fit any condition in herbal research domain.

Step 9 is identifying attribute or properties for each class. Result for this step is visualized by Figure 2. Figure 2 shows conceptual model for HREC's database. It consists of class, subclass, relationships, properties and related key for each class. The conceptual model will be used in physical design which in creating of HREC's database.

V. TESTING OF RESULT

This section discusses on rules created to test representation component assigned for conceptual model in this domain. This paper demonstrates rules for information retrieval in context of keyword searching. For example, user input: "*Extraction of Tongkat Ali using supercritical fluid extraction*". In order to simplify the view of the rules, eight terms mentioned above is are stated as array: term. For this example, author assumes that all terms or phrase input by user exists in herbal research ontology. We know that *extraction* is one of *process, Tongkat Ali* is one of *herb, supercritical fluid* is method of *extraction process.*

Table 4 provides rules for keyword searching input. There are several tables involved in retrieving information. Keyword searching which are table of *PREPOSITION*, *TERM*, *PREREQUISITE*, *TEMPORAL* and *MUTINC*. All mentioned table stored in domain ontology. First rule is identifying key concept. All prepositions have been identified and deleted. Rule 2 is checking available term which means to determine whether the word is in the domain or not. If exist, system carry id of the terms and go to



the next rule. Rule 3 is checking relationship between terms.

Figure 2. Conceptual model of HREC's database

This rule is only checking for the four extended component applied in this domain. It is important to evaluate whether the ontology designed is reliable or need to be refined. Outputs from this rule are id term, primary and foreign key of the terms. Next rule is checking property of data. Based on id term, primary key and foreign key of the terms, system could retrieve data from table of *HResearch, Herb,* and *Extraction*. Lastly, system display the information required. Figure 3 visualizes data flow model based on rules discussed above. Lines represent flow of the rules and dotted lines represent operation flow behind systems.

VI. CONCLUSION

This research discusses on how ontology technique can be used to represent conceptual model database design for herbal research domain. Four extended relationship components have been applied in the ontology representation. We found that the representation of the components is able to model information of herbal research domain. Hence, we can say that results in this paper support Sugumaran and Storey's approach. A conceptual model has been produced at the end of this phase. In order to ensure the conceptual model is well defined, a set of rules for keyword searching is created to verify preciseness of output produced. Although majority of scenario in herbal research domain can be represented by four extended representation, but some of scenario cannot be represented. Our future work will verify the database design using execution of rules for keyword searching through HREC's system.

ACKNOWLEDGMENT

We would especially like to thank project leader; Associate Professor Dr Azizah Abdul Rahman and members of project entitled, 'Information Architecture for Malaysian Natural Product Repository' funded by MOSTI. We would also like to thank to all academician and researchers who have helped us in this study.

REFERENCES

[1] M. Azlida, et al., "Proposal for Knowledge Capturing Tool for malaysian Herbs Reseracher," *Proc. The 4th postgraduate Annual Research Seminar (PARS'08)*, Universiti Teknologi Malaysia, 2008, pp. 185-188.J.

[2] A.P. Michelle, "Database Design and Management," *Getting It Right the First Time*, MOUNT VERNON DATA SYSTEMS LLC, May 2007.

[3] A. Borgida and R.J. Brachman, "Conceptual Modeling with Description Logics," *The description logic handbook: theory, implementation, and applications* Cambridge University Press 2003, pp. 359-381.

[4] V.C. Storey, "Comparing Relationship in Conceptual Modelling : Mapping to Semantic Classifications," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 11, 2005, pp. 11.

[5] A. Jashapara, *Knowledge Management;An Integrated Approach*, Prentice Hall, 2004, p. 93.

[6] K. Nigam and M. Mehrotra, "Comparative Analysis of Available Ontology Engineering Tools," *Proc. International Symposium on Information Technology 2008*, IEEE 2008, pp. 2582.

[7] T. Gruber, "Ontology (Computer scince)," *Book Ontology (Computer scince)*, Series Ontology (Computer scince), ed., Editor ed.^eds., Springer-Verlag, 2008, pp. 1-5.

[8] F.N. Natalya and L.M. Deborah, Ontology Development 101: A Guide to Creating Your First Ontology, Stanford University, 2000.

[9] M. Azlida and A.R. Azizah, "Organizing Herbal Research Knowledge Using Ontology Technique," *Proc. The 5th Postgraduate Annual Research Seminar (PARS'09)*, Universiti Teknologi Malaysia,, 2009-(In Press).

[10] V. Sugumaran and V.C. Storey, "The Role of Domain Ontologies in database Design: An Ontology Management and Conceptual Modelling Environment," vol. Vol 31, 2006, pp. 1064-1094.

[11] R.A. Aziz, et al., "The Need for Standardization in Large Scale Herbal Extraction," *The 4th traditional/Complementary Medicine Conference and Exhibition (INTRACOM 2002)*, 2002.

[12] N.M. Ahmad and M.R. Colomb, "Managing Ontologies: A Comparative Study of Ontology Servers," *Proc. ADC '07: Proceedings of the eighteenth conference on Australasian database*, Australian Computer Society, Inc, 2007.

[13] , "Overview of Herb Research," *Book Overview of Herb Research*, Series Overview of Herb Research, ed., Editor ed.^eds., Chemical Engineering Pilot Plant, 2008, pp.

TABLE IV. RULES FOR KEYWORD SEARCHING

	Extraction	*of*	*Tongkat*	*Ali*	*using*	*supercritical*	*fluid*	*extraction*
Phrase x =Assign term	[1]	[2]	[3]	[11]	[5]	[12]	[7]	[13]

Rule 1- Rule for identifying key concept

If ***: Phrase x is selected and***
Phrase x consists of multiple terms; term [x]
Term[2,5] is selected as preposition
Then ***: Term [2,5] are deleted***
Term [1,3,4,6,7,8] assigned as key concepts

Rule 2- Rule for checking available keywords

If ***: Term [1,3,4,6,7,8] is selected and***
Checking related-to, is_a, synonym, acronym,and hypernym.
There exists term [1,3,4,6,7,8] in domain ontologies
Term [1] and term [8] is similar,
Term [3] has "related to" relationship with term [4],and
Term [6] has "related to" relationship with term [7
Then ***: Either term [1] or term [8] is deleted.***
Combined term [3] or term [4] = term[9].
Combined term [6] or term [7] = term[10].
Get id terms for term [1,9,10]

Rule 3- rule for identifying relationship component

If ***: Id for Term [1,9,10] are selected and***
Term [9] has "prerequisite" relationship with term [1]
Term [10] has no relationship with of term [1,9]
Then ***: Id for term [1] is foreign key of term [9]***

Rule 4- Rule for checking property of data

If ***: Term [10] is selected and***
Term [10] is not property of term [9]
Term [10] is property of term [1]
Then ***: Get id for term [1] that have property = term[10]***
Get information from table HResearch
Display information that have id for term [1,9] and term[10] is property of term [1]

Figure 3. Data flow model for keyword searching