

A GENERALIZED APPROACH TO CONTENT CREATION USING KNOWLEDGE BASE SYSTEMS

By
K S Chudamani and H C Nagarathna
JRD Tata Memorial Library
IISc, Bangalore-12

ABSTRACT:

Library and information Institutions and centers all over the world are attempting to convert their holding into electronic storage. Collections of library document are accessed by users for their analytical and interpretational activities in research, academic consultation occupations and developmental work. *Quickly finding an article in the Microsoft Knowledge Base through search by using content keywords and query words is an essential requirement of the present era. Using the keywords and query words that are listed in one article may help find other articles that have similar content. However, some query words are only used for older content and may not help you find the most current information.* Protégé is a [free, open source](#) ontology editor and knowledge-base framework. The Protégé platform supports two main ways of modeling ontology's via the [Protégé-Frames](#) and [Protégé-OWL](#) editors. Protégé ontologies can be exported into a variety of formats. It works on the principles of using classes, slots, forms, instances, queries. Classes can have subclasses. Slots are the templates for each class through which data can be entered. Forms set the indexed instance set up. Instances are the data entry display forms using the defined slots. Queries are the search mechanisms through the metadata. Content creation is the ultimate unit of information. It has tremendous use in day-to-day life. It has significant role in decision making. The need of content management system started appearing effectively and efficiently since time immemorial. For creating an knowledge based system, the system designer has to select 4 basic elements. They are: Cards (Records), Thesaurus, Class no, Indexing. These components of the system are connected to the scope and application of the subject matter to which the content belongs to. The components are data about the content assertions. In this way it is easier to store, organize, process and retrieve the information. The Content Knowledge Base guided search allows users to find precise content technical information, Developer Centered articles, and information service. The institutional guided search functionality lets users fine-tune their queries based on established categories.

INTRODUCTION

Library and information Institutions and centers all over the world are attempting to convert their holding into electronic storage. Collections of library document are accessed by users for their analytical and interpretational activities in research, academic, consultation occupations and developmental work. One has to go through contents first and then accessing text on computer or manual system through linking of underlined words. Content creation from institution through simple to sophisticated indexing system plays a crucial role. Its bring together users of the emerging standards in this arena. *Quickly finding an article in the Microsoft Knowledge Base through search by using content keywords and query words is an essential requirement of the present era. Using the*

keywords and query words that are listed in one article may help find other articles that have similar content. However, some query words are only used for older content and may not help you find the most current information. This paper explores the use of protege2000 as a tool for content creation.

KNOWLEDGE-BASED SYSTEMS

Simple knowledge base system identifiers are communities of practice for developing standard vocabularies for specific local needs. User's access the contents of knowledge, the patterns they contain, textures and related layout and position information which allows users to query the system based on perceptual features.

The lack of standardized access and interchange formats impedes wider use of knowledge base systems resources. We have developed a demonstrator of Protege2000 (<http://protege.stanford.edu>) for the precise project that explored thesaurus-based query expansion in protégé.

Protégé is a “[free, open source](#) ontology editor and knowledge-base framework. The Protégé platform supports two main ways of modeling ontology's via the [Protégé-Frames](#) and [Protégé-OWL](#) editors. Protégé ontologies can be exported into a variety of formats. Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development. [\(more\)](#). Protégé is supported by a [strong community](#) of developers and academic, government and corporate users, who are using Protégé for knowledge solutions in areas as diverse as biomedicine, intelligence gathering, and corporate modeling. Protégé's flexible architecture makes it easy to configure and extend the tool. Protégé has an [open-source Java API](#) for the development of custom-tailored user interface components or arbitrary Semantic Web services”.

PROTÉGÉ 2000 AS A TOOL FOR CONTENT CREATION

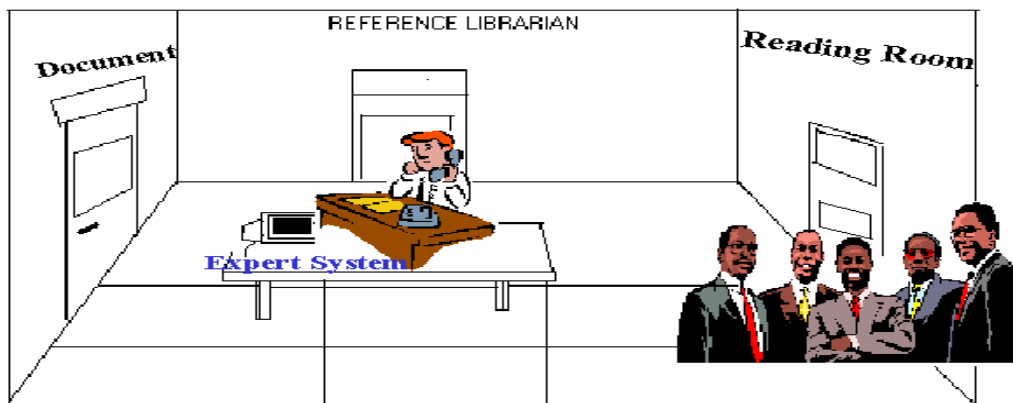
Protégé 2000 is a knowledge based system software. It works on the principles of using classes, slots, forms, instances, queries. Classes can have subclasses. Slots are the templates for each class through which data can be entered. Forms set the indexed instance set up. Instances are the data entry display forms using the defined slots. Queries are the search mechanisms based on the metadata.

For creating an knowledge based system, the system designer has to select 4 basic elements

They are:

1. Cards (Records)
2. Thesaurus
3. Class no
4. Indexing

C
a
r
d
s
t
h
e
s
a
u
r
u
s
C
l
a
s
s
N
u
m
b
e
r
g
h
i
x
e
d
i



These components of the system are connected to the scope and application of the subject matter to which the content belongs to. The components are data about the content assertions . In this way it is easier to store, organize, process and retrieve the information. The content is managed with the system that consists of database record file containing content components. Knowledge of context , not only makes information more meaningful, but it helps different individuals to use the same information at different times. Creating knowledge base system requires cards, thesaurus, Class no, indexing rules which are to be developed through an expert or artificial intelligent system to create concept based subject indexing strings of keywords covering semantics and syntax aspects. This knowledge based system that acts as a front end.

Expert systems provide techniques to perform design making tasks based on a programmed set of rules and logic with in specific subject areas.

Content creation is the ultimate unit of information. It has tremendous use in day-to-day life.

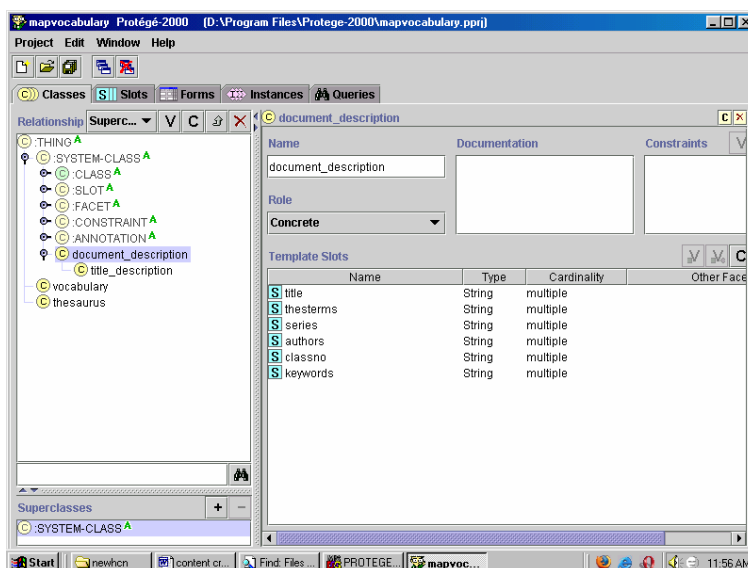
It has significant role in decision making. The need of content management system started appearing effectively and efficiently since time immemorial. Quality wise some content is conformed which is hence reliable, and the rest unconfirmed and therefore not much useful. It is available in several forms or types necessitating a content management system that reflect the changes that take place in the content. The system created at JRD Tata Memorial Library, IISc has three components, Namely,

1. Document description
2. vocabulary
3. thesaurus.

Application, as an initial experiment with the protégé, a rich client browser, displaying detail for thesaurus concepts, Controlled vocabularies, Document description, thesaurus have their own usefulness. Organizations continue to require efficient, effective access to controlled content for creating consistent data through metadata for their collections. All these are described in the following paragraphs.

Fig.1 is an description of the metadata used to store document description. Only six important metadata elements have been chosen to represent the documents. Each metadata is repeatable.

FIG 1



The metadata elements are

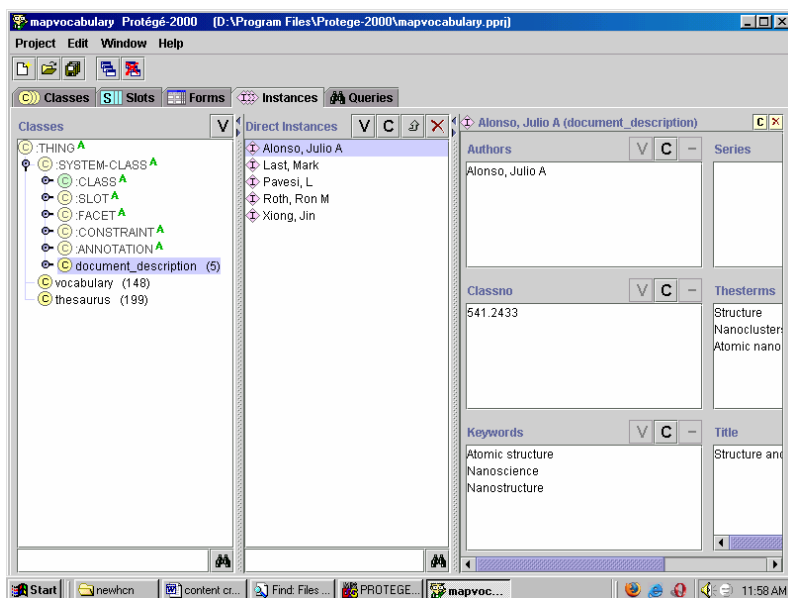
- Title
- Thes terms
- Series
- Authors
- Class No
- Keywords

Many controlled vocabularies are available for representing the content of documents and other resources. As applicable for each term retrieved, the above data is displayed in the research pane as expandable. The controlled vocabulary used here is the Library of Congress Subject Headings

The classification scheme used can be any scheme and multiple assignments can be made from the same scheme. In this context, for the project undertaken here which is a generalized approach to content creation multiple class numbers are given from DDC. Similarly thesaurus terms are also given from Library of Congress subject headings. This field is also repeatable. Also, the Instances can be viewed from a selected data element of the metadata.

Fig 2 is an display of the data entry form. The data entered are multiple. The screen opens out to accommodate multiple entries in the box. Data can be both numeric or text.

FIG 2



The protégé 2000 terms list is a vocabulary with some broader and narrower term relationships. We added more hierarchical structure to this list with concept form the documents to develop the metadata thesaurus. In the example above, the concept scheme moved from a term list to a thesaurus. Fig 3 is a display of the screen from the thesaurus maintenance system of the software. Its elements are the usual thesaurus metadata like lead in term, RT terms, BT terms, NT terms, etc.

FIG 3

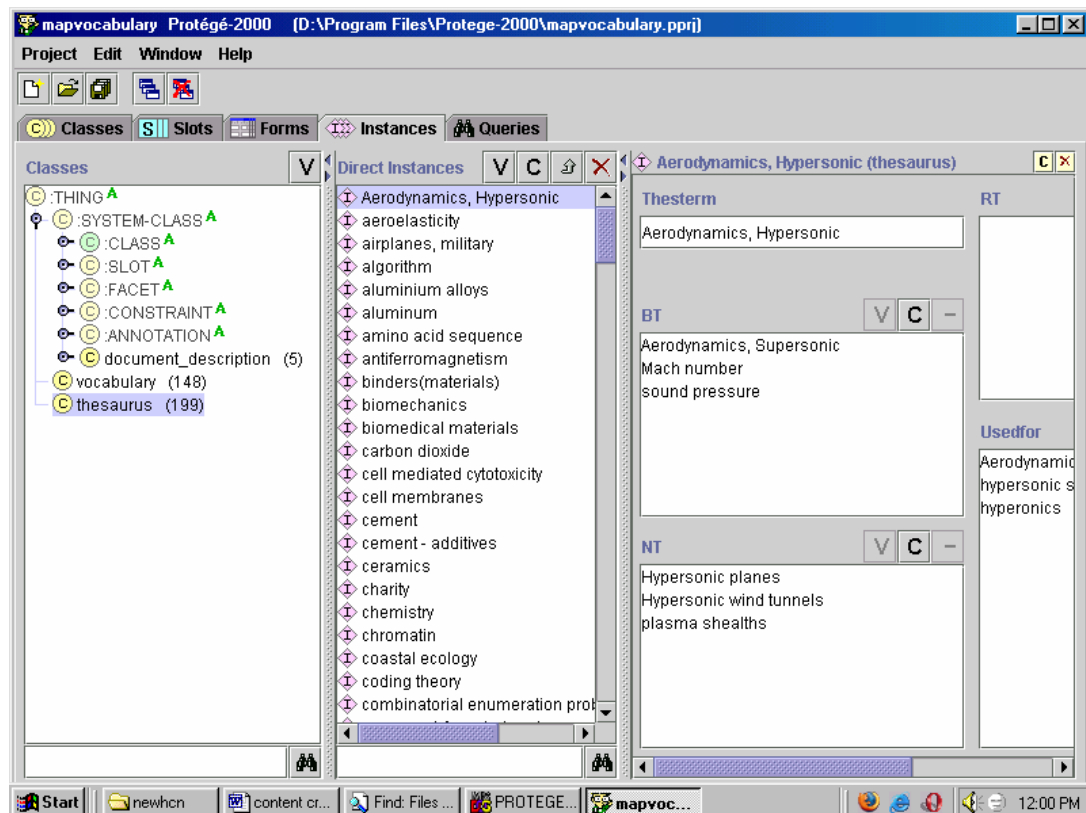
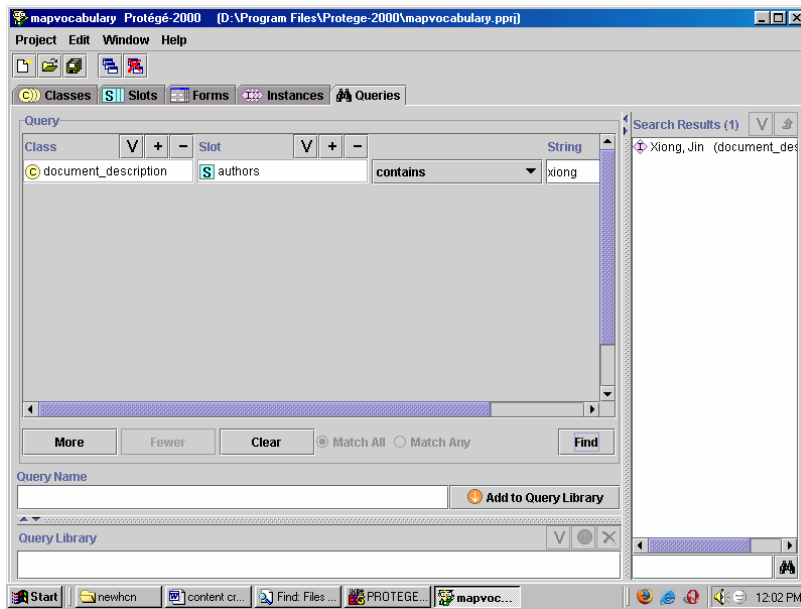


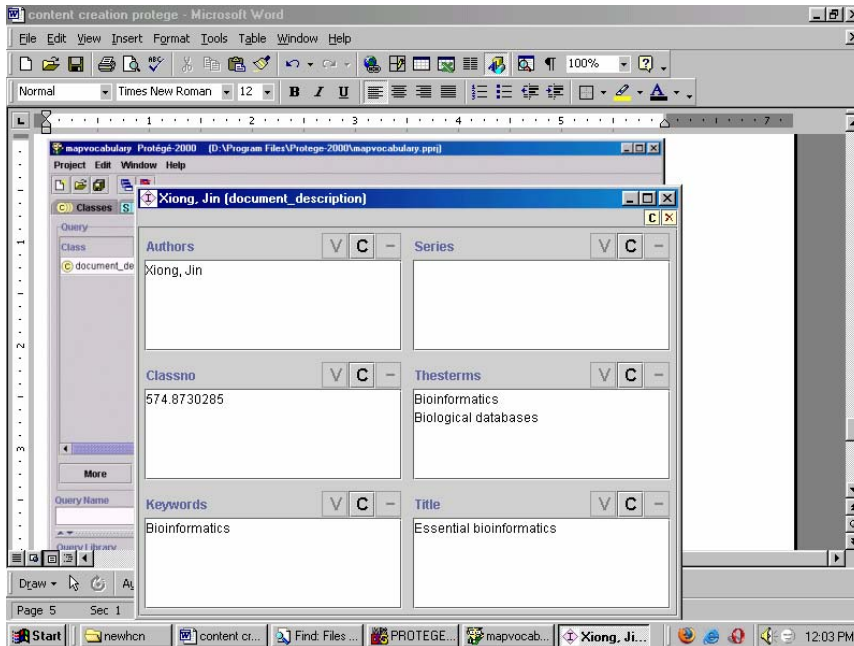
FIG 4 is a query module.. It has the ability to retrieve from any field and also any combination of fields. Each cell has the capacity to select the metadata and combine metadata for searching. It searches the given term anywhere in the specified fields.

FIG 4



Once the query is searched the result will be displayed as a record. By clicking on the record the full data in the metadata fields is displayed. This is presented in FIG 5

FIG 5



To enhance search capability a vocabulary is also created. This presents to the user the mapping of a common vocabulary through the thesaurus and also its class number. This is helpful as an entry point for the vocabulary to the user who is not aware of the system terminology. Fig 6 presents the data structure for this. It has the following elements:

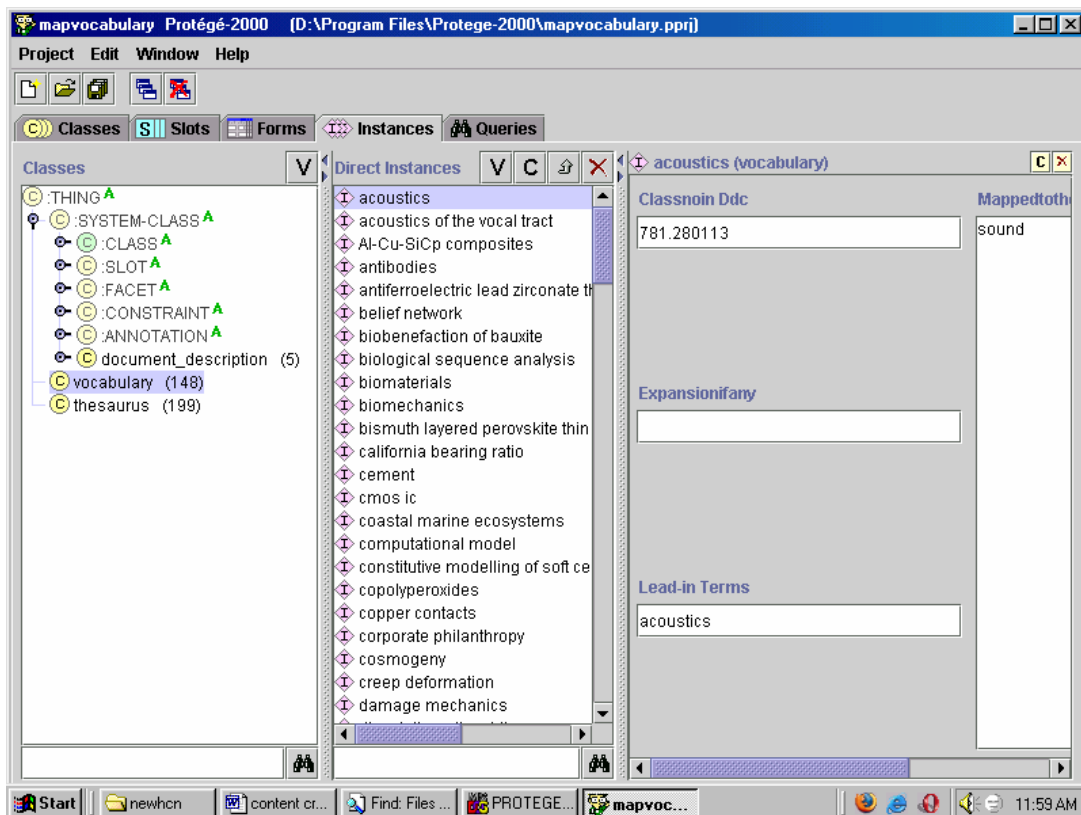
Lead-in terms

Classnoin Ddc

Expansionsifany

Mappedtothesaurus

FIG 6



This approach can be followed even in Greenstone digital library software and it can be web enabled. This approach will serve information searching with a more powerful and user friendly system. However, expanding the search requires interlinking of the databases appropriately.

Conclusion

The Content Knowledge Base guided search allows users to find precise content i.e. that is keyword or technical information Developer Centered articles, and information service. The institutional guided search functionality lets users fine-tune their queries based on established categories. Each search that a user performs automatically provides structured feedback to database administrator about the technical quality of the knowledge base content. This allows us to contribute new technical information continually to make the knowledge base and the guided search more effective.

The Content Knowledge Base guided search is a self-service user information search tool that allows searchers to access information quickly and efficiently from contents for precise information, Expert systems Develop content of articles books, and Knowledge base Management service, notes on policies and procedures. Each search that a user performs automatically provides structured feedback to information about the technical quality of the knowledge base. This allows us to contribute new technical information continually to make the knowledge base more effective..

Bibliography:

1. Information and knowledge management : by D Kamalavijayan, Macmillan: 2005
2. <http://protege.stanford.edu/plugins/owl/>
3. [http://protégé. Stanford.edu](http://protégé.Stanford.edu)
4. Information management in e-Libraries. Ed by S.Parthan and V.K.J.Jeevan.: Allied pub.
2002

