A Method for Developing a Domain Ontology: A Case Study for a Multidisciplinary Subject†

Sangeeta Deokattey,* Arashanipalai Neelameghan,** and Vijai Kumar***

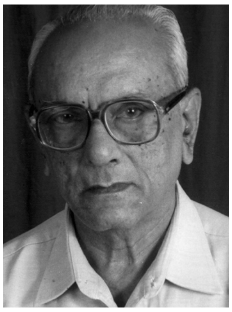
*Scientific Information Resource Divn., Bhabha Atomic Research Centre, Trombay, Mumbai-400 085,<sangeetadeokattey@hotmail.com>

**Documentation Research and Training Centre, Indian Statistical Institute, Bangalore-560 059,<anm2002@vsnl.net>

***Knowledge Management Group, Scientific Information Resource Divn., Bhabha Atomic Research Centre, Trombay, Mumbai-400 085,<vijai@barc.gov.in>

Sangeeta Deokattey has been working in the field of library-and-information science for the last 25 years. She developed and organized the Scientific and Technical Reports collection at the Small Enterprises National Documentation Centre. Later, at the Research Centre for Women's Studies, SNDT Women's University, she organized the special collection of Women's Studies research material and was one of the first in Mumbai to develop a database on Women's Studies. Her research interests include classification, indexing, and bibliometric analysis. She is presently working as a Senior Scientific Officer in the Scientific Information Resource Division, Bhabha Atomic Research Centre, India.

Arashanipalai Neelameghan retired as UNESCO/PGI Regional Adviser for Asia-Pacific. Former Professor and Head of ISI/DRTC, Bangalore. Special interests: knowledge organization (KO); KO tools; relationship among concepts; information communication with marginalized communities; education and training of information professionals.



Vijai Kumar is Associate Director, Knowledge Management Group and Head, Scientific Information Resource Division, Bhabha Atomic Research Centre, India. He is currently involved in executing a plan project for upgrading scientific and technical information resources, including setting up of a modern digital library. He is also a Senior Professor at the Homi Bhabha National Institute (a deemed University in Mumbai) and is associated with academic and scientific productivity analysis.

Deokattey, Sangeeta, Neelameghan, Arashanipalai, and Kumar, Vijai. **A Method for Developing a Domain Ontology: A Case Study for a Multidisciplinary Subject.** *Knowledge Organization, 37(3)*, 173-184. 11 references.

ABSTRACT: A method to develop a prototype domain ontology has been described. The domain selected for the study is Accelerator Driven Systems. This is a multidisciplinary and interdisciplinary subject comprising Nuclear Physics, Nuclear and Reactor Engineering, Reactor Fuels and Radioactive Waste Management. Since Accelerator Driven Systems is a vast topic, select areas in it were singled out for the study. Both qualitative and quantitative methods such as Content analysis, Facet analysis and Clustering were used, to develop the web-based model.

† The authors are grateful to Dr. Pitamber Singh and Dr. P.K. Nema, Nuclear Physics Division, BARC, Dr. S.B. Degweker, Theoretical Physics Division, BARC, Dr. R.C. Sethi (Retd.) Accelerator and Pulse Power Division, BARC for their expert comments, suggestions and advice.

1.0 Introduction

Ontologies as subjects of study cut across the domains of Philosophy, Logic, Artificial Intelligence, the Internet and Information Science. One of the earliest reported associations between ontologies and Information Science, was made by Dahlberg (1978). She established the link between cognitive processes, objective knowledge, concepts and their relations, and laid the ontological foundations of a modern classification system. The relationship between classification theory, cognitive science and artificial intelligence was demonstrated by Walker (1981) (Table 1). A modified version of the table displays how this relationship can be extended to ontologies too and shows the underlying unifying basis of all the four domains.

Both conventional knowledge organization tools such as classification schemes, thesauri and new tools (Denda 2005) such as ontologies, define concepts and relationships in a systematic manner. Thesauri (Lopez-Huertas 1997) and domain ontologies are both representational vocabularies restricted to a particular domain. The basic difference between them, lies in the use of descriptors and concepts to map a given domain. A few other differences have been elaborated in Table 2.

2.0 Current approaches for developing domain ontologies

Many of the current methods for building, merging and reusing ontologies have been developed by artificial intelligence researchers. Most of the current ontological projects use readily-available software tools to develop new ontologies. Ontolingua and Protégé are among the popular tools in use today. In the field of information science, research and development work on ontologies began around late 1990s. The Unified Medical Language System (UMLS) Metathesaurus (umls.info.nlm.nih.gov) is one of the most comprehensive domain ontologies in the field of biomedical information. Ding (2001) reviewed the importance of ontologies in the development of the Semantic Web. He discussed the definition of ontologies, kinds of ontologies, ontology tools, ontology language and a few ongoing ontology projects. Ding and Foo (2002) in another study, presented a two-part review. In the first part of the review, state-of-the-art techniques on semi-automatic and automatic ontology generation were detailed. The second part of the review dealt with ontology mapping and evolving. One of the first experiments to develop an ontology using a traditional knowledge organization tool was reported by Qin and Paling (2002). They used the controlled vocabulary of Education Resources Information Centre (ERIC) descriptors, to develop an ontology in the field of education, using Ontolingua. According to them, the major difference between a thesaurus and an ontology, lies in the values added through deeper semantics in describing digital objects, both conceptually and relationally.

3.0 The present study

This study uses a web-based approach to represent the ontology in the selected domain. For the purpose of the study, the following working definition of a domain ontology was adopted:

Cognitive Psychology	*Classification*	**Artificial Intelligence**	***Ontologies****
Thinking	Semantic linkages of elements	Heuristics and Predicate Calculus	*Identifying concepts and vocabulary of a domain*
Perception	Syntactic (metadata) Structure of the System and its rules	Production and Management Systems	*Developing the ontology language*
Remembering	Thematic relationships (Hierarchical or Associative)	Semantic Networks	*Conceptualization of the domain*
Imagination and Creativity	Facets of the corresponding elements of a Subject (based on Literary Warrant)	Frames	*Reusable modules of ontologies based on specific user requirements*

Table 1. Relationship between Classification theory, Cognitive Science, Artificial Intelligence and Ontologies. Source: Walker (1981) (*the table has been modified through the addition of the column on ontologies).

Thesaurus	Ontology
Rigid	Totally flexible
Unidimensional	Multidimensional
A thesaurus can either be manually created or machine-generated	An ontology can only be created using special software programmes.
Uses a controlled vocabulary, consisting of descriptors which represent the subject domain	Uses both free index terms as well as a controlled vocabulary in the form of concepts. A set of inference rules are also incorporated in an ontology that aid a user in discovering new knowledge
Controlled vocabulary terms are mostly restricted to domain knowledge	An ontology incorporates both domain knowledge and operational knowledge
Built on the concept of Literary Warrant. Terminological concepts and their relationships (including incorporation of new terms) are purely based on Literary Warrant.	Literary Warrant as well as other problem solving practical tools, devices and elements. Terminological concepts as well as other characteristics and properties are defined in the purpose of the ontology and may be shifted to new slots, or deleted from existing slots, depending on emerging applications.
Only valid descriptors are used both during storage as well as retrieval	Though a vocabulary is an inherent part of an ontology, other concepts and terms are freely used, both during storage and retrieval, as well as for other applications
There are only three kinds of relationships between any two descriptors in a thesaurus; Hierarchical, Equivalence and Associative	The relationships between classes or properties are potentially polyhierarchical; a class may be a subclass of one, two, three or more superordinate classes
The superordinate and subordinate classes are determined solely by the domain	The superordination and the subordination of classes is mostly decided by the ontology developer depending on the purpose of the ontology.
Descriptors in a thesaurus exist purely at a conceptual level.	Terms or classes in an ontology can be assigned values.

Table 2. A few differences between a thesaurus and a domain ontology

> A domain ontology is a knowledge organization tool on a specific subject domain either pure, interdisciplinary or multidisciplinary and which incorporates both a metadata schema and a vocabulary. It is based on WWW standard for metadata encoding. It uses both available controlled vocabulary terms and free index terms and is based on the concept of literary warrant.

3.1 *The multidisciplinary subject domain: accelerator-driven-systems (ADSs)*

Accelerator-driven-systems is a narrow subject domain, a part of the field of nuclear science and technology. The domain of nuclear science and technology is itself multidisciplinary, as it encompasses physics, chemistry, engineering and biology disciplines. It is also interdisciplinary as it is an amalgamation of all these domains. Accelerator-driven-systems is a part of the broad area of physics, especially particle physics and atomic and nuclear physics. It also includes nuclear and reactor engineering, reactor fuels and radioactive waste management. Basically, an accelerator driven system is a hybrid system. It combines a particle accelerator, particularly a high power proton accelerator with a subcritical nuclear reactor to accomplish various tasks. Research and development work on accelerator driven systems gains importance due to their inherent safety features, their use of thorium as fuel, and their non-proliferative advantage (Carminati et al 1993; Gokhale et al 2006). In this study, a domain ontology was developed on accelerator driven systems. Though there are several applications of accelerator driven systems, the study was restricted to information on a) accelerator driven systems in general, b) accelerator driven systems for energy production (energy amplifier (EA)) and c) accelerator driven systems for transmutation of waste (accelerator transmutation of waste (ATW)) applications.

3.2 *The method used for the present study*

There are three methodological options (Mai Chan and Lei-Zeng 2002) for developing new systems or tools for knowledge organization purposes, especially in networked environments:

- translation, i.e., to integrate one or more existing tools or systems to create more powerful ones;
- merging, i.e., mapping of a vocabulary from a larger system and developing a new one or mapping two or more existing schemes;
- creating from scratch, i.e., developing an entirely new system.

In this case study, the mapping/merging system was used and a new knowledge organization tool was developed using an existing larger tool; the International Nuclear Information System (INIS) thesaurus. This is more economical and it also ensures interoperability, since the INIS thesaurus is a standard vocabulary control tool, in the field of nuclear science and technology.

Appropriate bibliographic searches conducted on the INIS database (under the guidance of subject experts), yielded a total of 2,238 bibliographic records on ADSs in general: 694 records on ATW and 129 records on EA. Figure 1 shows the query formulation. All the descriptors (both computer-assigned and indexer-assigned) from each bibliographic record were downloaded separately. Free-text keywords were also selected from the abstracts of the downloaded records. Using the basic principles of content analysis, facet analysis and clustering, an ontological framework for ADSs was developed.

#20 (accelerator driven energy production) or (energy amplifier and accelerator) or (ADEP and accelerator)
#19 ((accelerator driven*) in DE) or (ADS and accelerator) or (accelerator driven systems) or (AAA and accelerator) or (advanced accelerator applications) or (hybrid reactor and accelerator) or (accelerator driven subcritical reactor) or (ADSS and accelerator) or (ADSR and accelerator)
#18 (accelerator transmutation of waste) or (accelerator driven transmutation technolog*) or (ADTT and accelerator) or (ATW and accelerator)
#17 ATW (Accelerator Transmutation of Waste) and accelerator
#16 ADEP (Accelerator Driven Energy Production) and accelerator
#15 ADTT (Accelerator Driven Transmutation Technologies) and accelerator
#14 ADSR (Accelerator Driven Subcritical Reactor) and accelerator
#13 ADSS (Accelerator Driven Subcritical Systems) and accelerator
#12 accelerator driven subcritical reactor
#11 hybrid reactor and accelerator
#10 thorium breeding
#9 energy amplifier and accelerator
#8 advanced accelerator applications
#7 AAA (Advanced Accelerator Applications) and accelerator
#6 accelerator driven transmutation technolog*
#5 accelerator driven energy production
#4 accelerator driven systems
#3 ADS (Accelerator Driven System) and accelerator
#2 (accelerator driven*) in DE
#1 accelerator transmutation of waste

Figure 1. Query formulation used for searching the INIS database

The following steps formed part of the method:

1. Downloading of the assigned descriptors (a total of 4808) and identification of new free-text keywords on ADSs (a total of 474), from the abstracts of the downloaded records, using content analysis.
2. Numerical listing of the descriptors and free-text keywords on ADSs, based on frequency count and selection of those with a threshold level of more than 2. A total of 3085 descriptors and all the 474 keywords were finalized.
3. Developing a semantic network of selected descriptors and keywords on ADSs through the creation of *Excel* files. All the 3085 descriptors and the 474 keywords were linked to each other through subordinate, superordinate and affinitive relationships. Three separate files were developed on ADSs, ADT and EA subfields. A sample of one of the *Excel* files is shown in Figure 2.
4. Developing clusters on the basis of frequency count of the descriptors. Descriptors with very high frequency counts were identified as core descriptors or concepts and clusters were developed around them, through the generation of one-to-many correspondences, using facet analysis. (For example, for the descriptor 'targets,' which forms part of the larger cluster 'accelerators,' five facets were identified: type of targets; purpose or application of the target; parts of the target; specific targets; and material of the target. The material facet was again subdivided into two parts: solid material such as metal and liquid material such as a

Descriptor/Keyword	Superordinate	Subordinate	Affinitive
absorption-;			heterogeneous-effects; radiations; self-shielding; shielding; slowing-down;stopping-power; transmission
absorption-spectroscopy;	spectroscopy-		infrared-spectra;
abstracts-;	document-types;	leading-abstract;	
acceleration-;			accelerators-; velocity
accelerator-breeders;			accelerators-; breeder-reactors; breeding; fissionable-materials;nuclear-fuels
accelerator-driven-transmutation;	transmutation-		accelerator-breeders;accelerators; radioactive-waste-processing
accelerator-facilities;		target-chambers	accelerators-; beam-dumps; beam-monitors; laboratory-equipment; reaction-product-transport-systems
accelerators-		coherent-accelerators; collective-accelerators; cyclic-accelerators; electrostatic-accelerators; heavy-ion-accelerators; linear-accelerators; meson-factories; particle-beam-fusion-accelerator	acceleration; accelerator-breeders; accelerator-driven-transmutation; accelerator-facilities; beam-dumps; beam-dynamics; beam-separators; impact-fusion-drivers; isotope-production; particle-boosters; storage-rings; targets; target-chambers
accident-insurance;	insurance-;		accidents-;

Figure 2. A sample of the interlinked EXCEL file of keywords / descriptors on Accelerator Driven Systems

liquid combination of lead and bismuth at high temperatures). The highest ranking core descriptors were: fuels, transmutation, accelerators, reactors, elements, materials, nuclear-reactions, subcritical-assemblies, and management. A total of nine separate *Excel* files on all these nine core descriptors were developed. Thus the final set of 12 interlinked *Excel* files formed a semantic network of the entire field of accelerator-driven-systems. Certain descriptors could not be slotted into any of the nine major clusters and these were just listed alphabetically. For example:

Data: data-covariances, evaluated-data, experimental-data, theoretical-data

Calculation methods: composite-models, mathematical-models, Monte-Carlo-method, numerical-analysis, simulation

Research programs: coordinated-research-programs

International organizations: national organizations, French organizations, Japanese organizations.

5. Uploading the above semantic network onto a three tier, web-based architecture of active server pages, Web server (IIS), and MS Access.
6. Saving the other bibliographic details of the downloaded records on ADSs, such as author, title etc. as html files and uploading them onto the web-based platform.
7. Developing a home page and a user interface for the prototype domain ontology on ADSs, with security provisions.
8. Developing separate hyperlinks to the three select areas of ADSs on the home page.

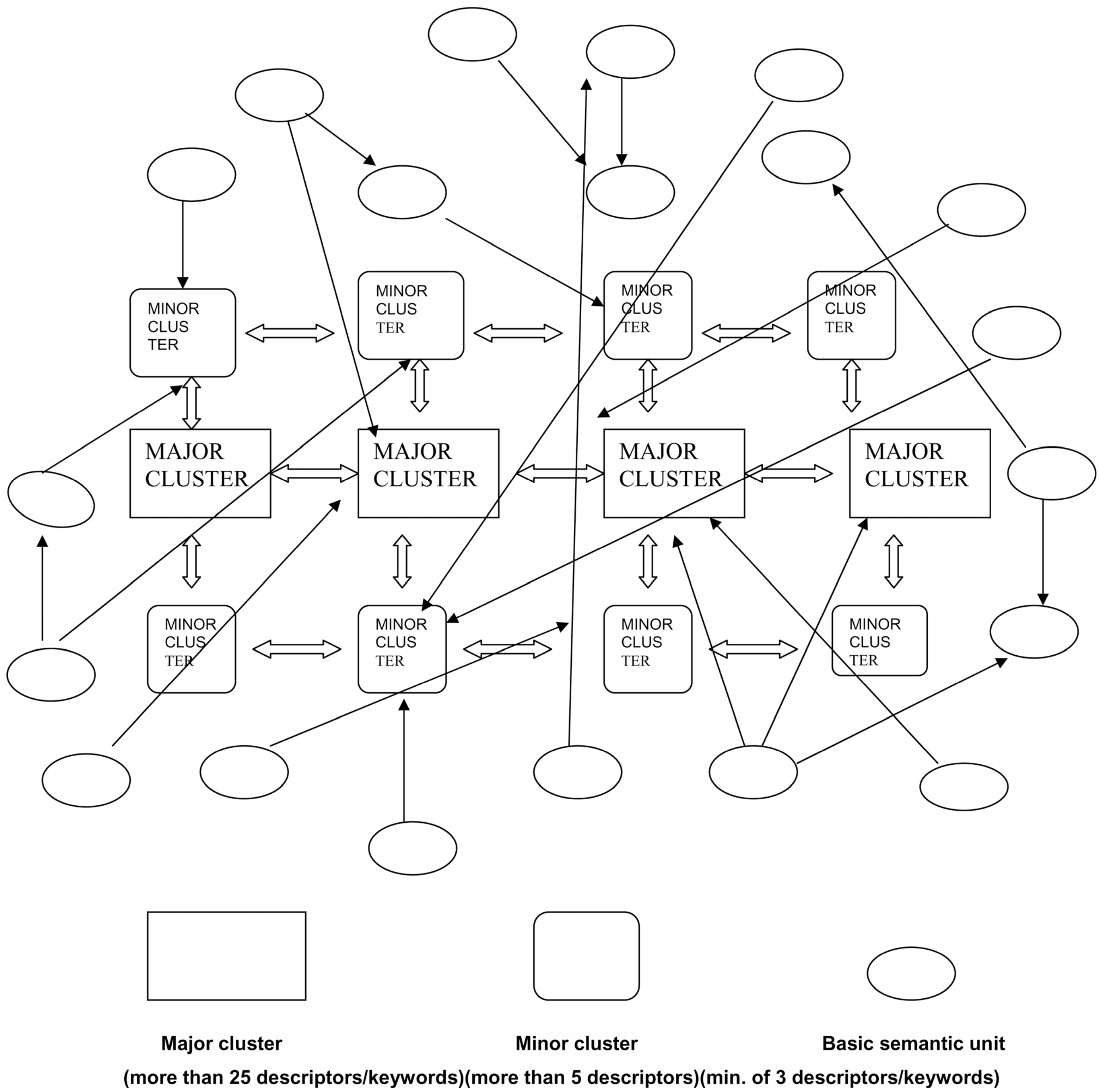


Figure 3. Prototype model of the domain ontology on Accelerator Driven Systems: a graphical representation

9. Testing and validation of the prototype model by experts on ADSs. (A prototype model of the domain ontology can be seen in Figure 3).

3.3 The domain ontology on ADSs as a knowledge organization tool

The author, title, abstract and keywords/concepts fields were made searchable. A user can search this domain ontology through the use of alphabetical A-Z drop-down menu of keywords/concepts; selecting a particular alphabet, displays all the keywords/concepts under that alphabet. There is a provision to add even full-text information on ADSs to the domain ontology and make it searchable.

The URL of the domain ontology is http://brew-2006.com/barc. After typing in the LOGIN ID and PASSWORD, the homepage (Figure 4) is displayed. Figure 5 illustrates the term "ACCELERATOR DESIGN CONCEPTS." This term, suggested by ADS experts, brings all the descriptors and keywords under one umbrella. The user can see at a glance, all the re-

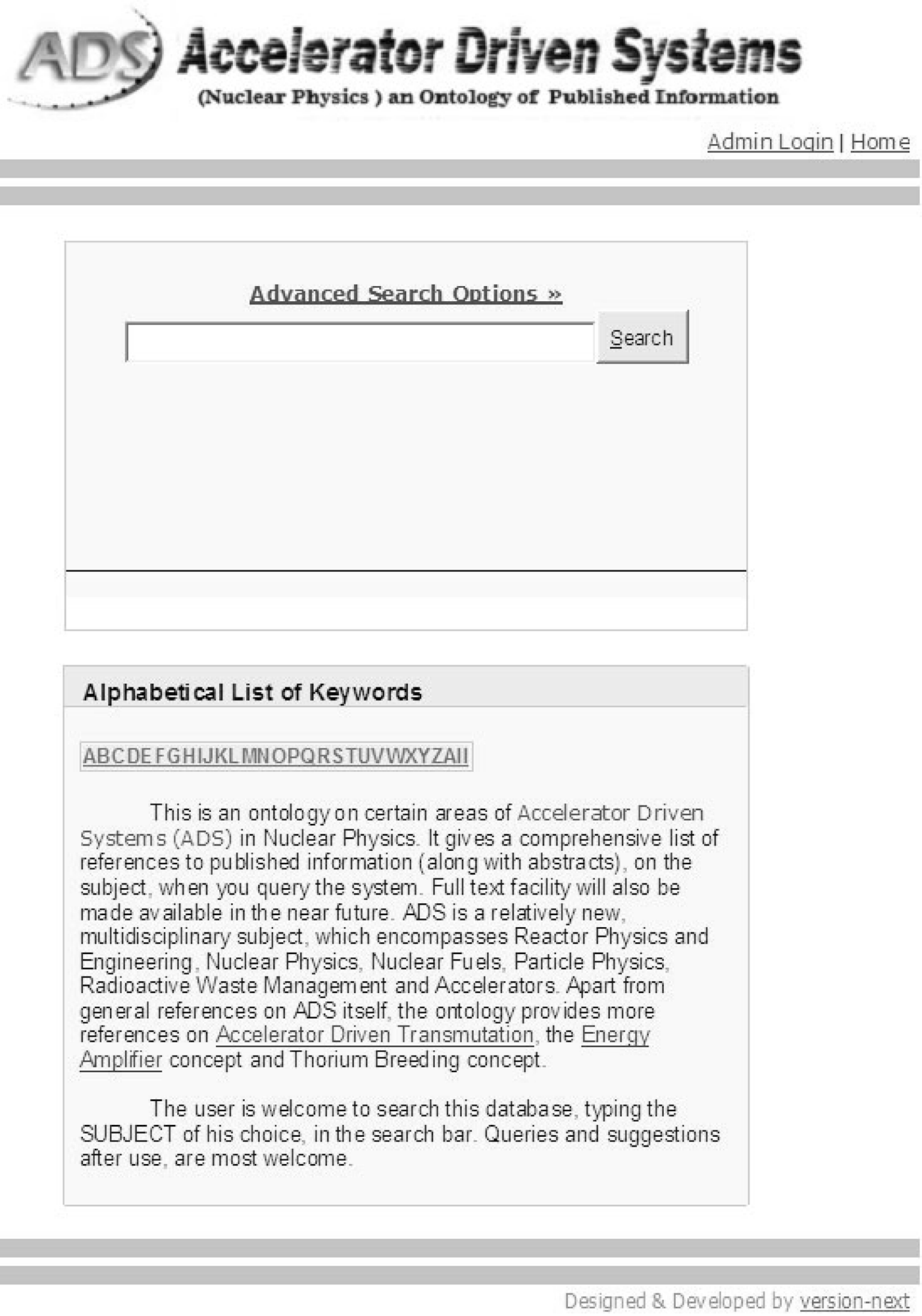


Figure 4. Homepage of the domain ontology on ADS

lated keywords associated with it and select any or all of the keywords to include them in his/her search. The "advanced search options" can be used to conduct searches other than on subject. Apart from the assigned descriptors and the additional free text keywords selected from the abstracts, ADS experts suggested the addition of the following set of twenty one new terms, which they thought were important. These were also incorporated into the domain ontology.

ADS Accelerator Driven Systems
(Nuclear Physics) an Ontology of Published Information

Back : Close : Home

ACCELERATOR DESIGN CONCEPTS

Keyword	☑ ACCELERATOR DESIGN CONCEPTS
Types	☐ AAA ☐ ABC ☐ ABR ☐ AD-RCNPS ☐ ACS ☐ ACWL ☐ ADAPT ☐ AD-GT-MHR ☐ AD-SPR ☐ ADC ☐ ADTT ☐ ADT ☐ ADTF ☐ ADMAB ☐ ADONIS ☐ ADPT ☐ ADR ☐ ADSR ☐ AMSB ☐ AMSTER ☐ APDF ☐ APT ☐ ARFR ☐ ATR ☐ BTA ☐ CAPRA/CADRA ☐ CRC ☐ CRIEC ☐ CSMSR ☐ CYRIC ☐ ENF ☐ ENR ☐ EDLE ☐ ETA ☐ F-EA ☐ FEAT ☐ FNS ☐ GEDEON ☐ HADRON

Figure 5 (continued on next page)

- [ ] HGF
- [ ] HILBILAC
- [ ] HYPER
- [ ] INDUS-1
- [ ] IPHI
- [ ] IPNS
- [ ] ISAAC
- [ ] ISIS
- [ ] ISNS
- [ ] ISOLDE
- [ ] ISSC
- [ ] JHF
- [ ] J-PARC
- [ ] KALLA
- [ ] KENS
- [ ] Kharkov-Linac
- [ ] KOCOS
- [ ] KOMAC
- [ ] KUCA
- [ ] KURRI
- [ ] LADR
- [ ] LAMPF
- [ ] LANSCE
- [ ] LASREF
- [ ] LEDA
- [ ] LEHIPA
- [ ] LPSS
- [x] MASURCA
- [ ] MEA
- [ ] MEGAPIE
- [ ] MINERVE
- [ ] MLNSC
- [ ] MMF
- [ ] MMKFK
- [ ] MRTI Design
- [ ] MSNES
- [ ] MSRE
- [ ] MUSE
- [ ] MYRRHA
- [ ] NSNS
- [ ] NSP
- [ ] n-TOF
- [ ] OMEGA
- [ ] Phasotron
- [ ] PHENIX
- [ ] PHOENIX
- [ ] PIREX
- [ ] PUSCA
- [ ] RAPSODIE
- [ ] RAS

Figure 5 (continued on next page)

	RCS RF-Linac RIA SATURNE SCANDAL SCNES SILHI Singletron-accelerator SINQ SLAC SNS SOC SPIN STF Superphenix Synchrophasotron TARC TASSE TEF Tevatron-collider TIARA TMT TRASCO TRISPAL TRITRON TRIUMPF XADS YALINA
General-design	RAJPA
Target-design	windowless-target
Blanket-design	lead-channel-blanket-design
Target-material	Tantalum Tin Tungsten
	Show Result

Designed & Developed by version-next

Figure 5. A display of the term “ACCELERATOR DRIVEN CONCEPTS”

1. Medium Energy Beam Transport (MEBT)
2. Quality factor
3. Power coupler
4. Twiss parameter
5. Waveguides
6. Permanent magnet quadrupole
7. Tuners
8. Postcouplers
9. Dipole stabilizer rods
10. Beam diagnostics
11. FOcusing and DefOcusing (FODO)
12. FOcusing- FOcusing and DefOcusing-DefOcusing (FOFODODO)
13. Separated Drift Tube Linac (STDL)

14. Coupled Cavity Linac (CCL)
15. Shunt impedance
16. Quadrupole gradient
17. Phase advance
18. Electron trap
19. Coupling cells
20. End cells and
21. Circulators.

3.4 Maintaining and updating the prototype domain ontology on ADSs

Accelerator-driven-systems is an emerging and widely researched subject domain. New keywords, which would be an integral part of the growing vocabulary on ADSs can be added to the domain ontology. New linkages can also be added to the existing keywords or they could be modified altogether. This is an ongoing process and can be done under expert advice and guidance.

4.0 Conclusions

Conceptualization is at the core of a domain ontology. To identify concepts in a narrow domain, (the domain of ADSs itself is comparatively new (early 1990s)) descriptors in a thesaurus can be used as basic building blocks and to harness these descriptors, appropriate query formulation is essential. Thus in the present study, descriptors from the INIS thesaurus were used, to organize the growing and vast amount of information on ADSs and a suitable knowledge organization tool was developed. International databases either bibliographic or full-text are structured and contain fields such as descriptors and abstracts, which are rich in vocabulary. The use of free-text keywords from the abstracts became necessary as a number of important terms were identified by the subject experts, which were not part of the INIS vocabulary. The prototype domain ontology was developed using a standard method, based on information science theory and practice. This tool was easy to develop, made use of available resources of information, was web-enabled, flexible and easy-to-use and more importantly, it was able to semantically map a growing multidisciplinary and interdisciplinary subject area: accelerator-driven-systems. It was also able to handle both the elements which are an integral part of any knowledge organization tool; vocabulary as well as metadata elements. The domain ontology developed in this study, is thus a dual-purpose knowledge organization tool and is interoperable in any digital environment. The data files can also be saved as XML files. It is easy to update as it uses terms both from a standard vocabulary as well as free text terms and is therefore very flexible. A flat structure for a domain ontology as displayed in Figure 4 was found to be more appropriate for an interdisciplinary subject domain such as accelerator-driven-systems, instead of the traditional "top down" or "bottom up" approach. The same structure can be applied to any other interdisciplinary and multidisciplinary subject domain. (The authors conducted a small pilot study on the interdisciplinary domain of biotechnology, using the same method). A domain ontology enriches the quality of semantic information in any given domain due to the various linkages provided between the keywords, through the identification and representation of one-to-many correspondences. These correspondences in the form of lateral or RT relationships have to be precisely identified and appropriately represented. In the present study, the identification of these relationships was done under the guidance of subject experts; the representation of these relationships was achieved through facet analysis and the type of relationship existing between any two keywords or group of keywords, was exactly specified to aid both the novice user as well as to save the time of the subject expert. Clustering of descriptors into concepts (on the basis of users' requirements) was seen to enhance retrieval of relevant information and was found to be a useful method for more effective knowledge organization in interdisciplinary and multidisciplinary subject domains.

References

Carminati, F. et al. 1993. An energy amplifier for cleaner and inexhaustible nuclear energy production driven by a particle beam accelerator. Geneva, CERN (CERN-AT-93-47).

Dahlberg, Ingetraut. 1978. Ontical Structures and Universal Classification. Bangalore, Sarada Ranganathan Endowment for Library Science.

Denda Kayo. 2005. Beyond subject headings; a structured retrieval tool for interdisciplinary fields. *Library resources & technical services* 49: 266-75.

Ding, Ying. 2001. A review of ontologies with the Semantic Web in view. *Journal of information science* 27: 377-84.

Ding, Ying and Foo, Schubert. 2002. Ontology research and development. Part 1: a review of ontology generation. *Journal of information science* 28: 123-36.

Ding, Ying and Foo, Schubert. 2002. Ontology research and development. Part 2: a review of ontology mapping and evolving. *Journal of information science* 28: 375-88.

Gokhale, Pratibha A., Deokattey, Sangeeta, and Kumar, Vijai. 2006. Accelerator driven systems (ADS) for energy production and waste transmutation; international trends in R&D. *Progress in nuclear energy* 48n2: 91-102.

López-Huertas, María J. 1997. Thesaurus structure design: a conceptual approach to improving interactions. *Journal of documentation* 52: 139-72.

Mai Chan, Lois and Lei-Zeng, Marcia. 2002. *Ensuring interoperability among subject vocabularies and knowledge organization schemes: a methodological analysis.* 68th IFLA Council and General Conference, Aug. 18-24.

Qin, Jian and Paling, Stephen. 2001. Converting a controlled vocabulary into an ontology: the case of GEM. *Information Research* 6n2. Available at http://InformationR.net/ir/6-2/paper94.html.

Walker, Donald E. 1981. The organization and use of information; contributions of information science, computational linguistics and artificial intelligence. *Journal of the American Society for Information Science* 32: 347-63.