Semantic Network Based Modelling to Design a Knowledge-Driven Decision Support Tool for Floral Diversity

Sudhansu Sekhar Dash
Botanical Survey of India (BSI), Kolkata, India
Email: ssdash2002@yahoo.co.in

Mina Desai, Arghya Ghosh, and Samaresh Das
Centre for Development of Advanced Computing (C-DAC), Kolkata, India
Email: {mina.desai, arghya.ghosh, samaresh.das}@cdac.in

***Abstract*—This paper intends to design and simulate a broad knowledge-driven Decision Support Tool (DST) for the floral diversity domain by applying semantic network modelling theory. It emphasizes on how an applied discipline of Information Systems and Technologies uses knowledge gathered over hundreds of years by the scientists and researchers of Bio-diversity domain to arrive at a broad knowledge-driven DST. It introduces the theory of semantic network, concept map modelling and the method of building knowledge-based repository from the vast expanse of research, study and publication knowledge to arrive at a framework for knowledge-driven DST. The said modelling is explained in the light of floral diversity. The system architecture is described with the design aspects of knowledge-driven DST in line with generic operations that can be performed with it. The paper throws light on pros and cons of the designed model. It concludes with the implementation details to show how the field of computerized DST is expanding by using emerging technologies and to create new application for Bio-diversity discipline.**

***Index Terms*—concept map, DST, floral diversity, knowledge-driven tool, semantic network**

I. Introduction

Knowledge-driven Decision Support Tool (DST) has been a popular concept since the emergence of Information Technology [1]. Its application in any specific domain of any discipline offers decision making capabilities based on structuring of textual information. The biodiversity community is no exception. Biologists and decision makers have been involved in relentless research generating huge databases on the species; its status, habitats, socio-economic impacts, threats to its existence etc. Such textual information gives scientific basis for decision-making. The present day technological advances in computing have made us available with the spatial data as well as Geographical Information System. There is a paradigm shift in the way the knowledge base is used by application of Artificial Intelligence to come up with crucial decision support system. The decision involving floral research issues, invariably considers geographical parameters such as location (latitude and longitude), distance, direction, proximity adjacency, topography etc. along with textual data. The textual data may be in the unstructured format that needs to be structured by application of semantic network.

Manuscript received September 26, 2014; revised February 10, 2015.

II. Objective

A great deal of progress has been made in developing and refining the techniques of the study of floral diversity of the world. The primary objective of this paper is to throw light on how the decisions are drawn from the knowledge gathered over hundreds of years of researches by the scientists of Bio-diversity domain using Knowledge-driven Decision Support Tool (DST). The goal of this tool is to arrive at decisions about Flora of India based on varied attributes such as taxonomical classification, distribution, publication and research data. For example, with the help of detailed taxonomic study, distribution of a species in a particular region or researchers working in a particular field or particular group, a crucial research oriented decision can be derived by implementing decision support systems in the field of floristic diversity. Knowledge representation and structuring of textual data can be done using Semantic Network Model. The model is basically a frame network which represents concepts and semantic relations existing between them. This knowledge can be manipulated by the DST to infer decisions [2]-[4].

III. Semantic Network Modelling

By definition a semantic network, or frame network, is a network which represents semantic relations between concepts [5]-[7]. This is often used as a form of knowledge representation. It is a directed or undirected graph consisting of vertices, which represent concepts, and edges, which represents relation between concepts.

doi: 10.12720/jomb.5.1.53-57

This approach to represent knowledge has facilitated the model to classify the flora hierarchically in order to achieve drill-down modelling from the root class, i.e., group to the ultimate leaf class, i.e., floral species.

A. Concept

Novak (1984), based on Ausubel's (1968; 2000) and Toulmin's (1972) work, defines "concept" as a perceived regularity or pattern in events or objects, or records of events or objects, designated by a label.

B. Concept Maps

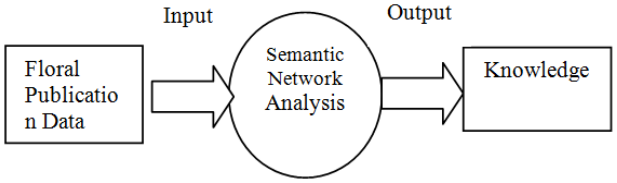
Concept maps are graphical tools for organizing and representing knowledge. They include concepts, usually enclosed in circles or boxes of some type, and relationships between concepts indicated by a connecting line linking two concepts [8], [9]. Words on the line, referred to as *linking words* or *linking phrases*, specify the relationship between the two concepts.

Underlying every semantic network is a graph, consisting of a set of nodes (also called vertices) and a set of edges or arcs that join pairs of nodes. An edge is an undirected link between two nodes, and a graph containing only edges is said to be undirected. An arc is a directed link between two nodes, and a graph containing only arcs is said to be directed.

The concepts are represented by nodes. The nodes having relationships are interlinked to form a tree like structure. We can navigate through the nodes based on the required criteria and relationships to obtain a particular result. In our case it primarily consists of floral database of the published Flora of India volumes.

The basic steps involved in the process are:

- To understand the semantics of the floral diversity and model them to arrive at a semantic network diagram having concepts and relations.
- For each concept the floral research and publication data, map it to relational database with hierarchical drill down of tuples. Hence, a structured database is achieved out of publication data.



- A web based drill down search and retrieval system is developed for dissemination of the above knowledge.

Decision can be taken based on the following parameters:

- Author-wise
  - Taxon-wise
  - Distribution Analysis
  - Period-wise
  - Classification

The unstructured floral taxon data available from Flora of India publication has been structured by identifying concepts and relations and put them into network diagram [10]. The network frames a relationship among the concepts.

For instance, we can find from the diagram that there are publications, books or journals which consists of unstructured knowledge about Taxon. Taxon has some attributes like Distribution, Illustrations, Synonyms etc. and has relation with Taxon Master, *citation* etc. There is a relation between a *taxon* and its *distribution*, *citation*, *taxon master* etc. The relationship between a *taxon* and *distribution* is 'found in' and that of *citation* is 'has'. Likewise, we can find the relation between *taxon* and its *type*, viz., genus, species etc. by the relation between *Taxon* and *Taxon Master* and between *Taxon Master* and *Type*, e.g., *Taxon* 'included in' *Taxon Master* 'is of' *Type*. This network allows us to reach to the leaf concept from the root concept by taking a suitable root and understanding the relations. Thus the knowledge extraction becomes fast, simple and easy.

Fig. 1 depicts the semantic diagram representing floral taxon.

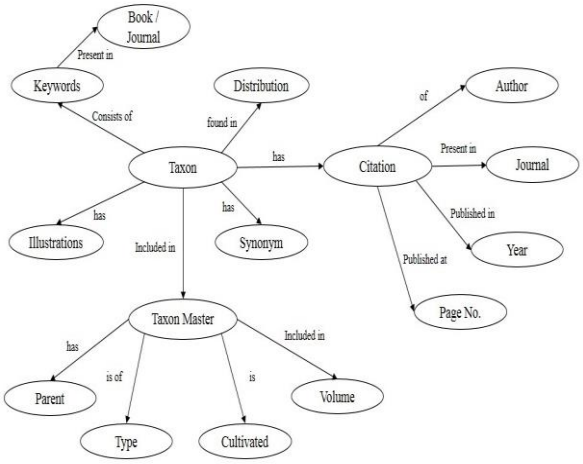


Figure 1. Semantic diagram representing floral taxon

IV. Knowledge-Driven Decision Support Tool (DST) in the Light of Floral Diversity

A. Attributes of the DST:

- DST is interactive and provides ease of use.
- Adaptable.
- Supports modelling and analysis.
- Data access.
- Highly flexible and user friendly.
- Decision makers in semi structured and unstructured problems.

Decision Support Tool (DST) is an application of Herbert Simon Model, which has the following phases:

1) Intelligence phase

It searches for the conditions that call for decision. In light of Bio diversity domain, the taxon distribution, diversity, density, research need etc. call for decision making.

2) Design phase

It designs possible routes and alternative actions of solutions. In our case, DST has been so designed as to get the best possible route to leaf node.

3) Choice phase

It deals with selecting the best course of route for drilling down to the leaf node.

4) Implementation phase

Adopting the selected course of action in decisive situation with human intervention.

In the light of Bio diversity domain, DST helps in performance analysis and decision making. DST can be built around the knowledge generated out of publication.

B. The Process Flow of the Application is as Follows:

1) Collection of bio-diversity data.
2) Data is ingested via the system in a pre-defined method.
3) DST processes using various geographical and other parameters.
4) It performs various analysis, based on input and available data, to generate new information.
5) Internally it uses semantic network analysis for decision making.
6) It captures the decision making rules and filters voluminous information based on input and decision making criteria.
7) The resultant information is made available to the user.

V. Features

The web-based application developed based on the semantic network modelling as described above has the following features:

1) The web application will facilitate to get information regarding flora by Family, Genus and Scientific Name.
2) It provides Family wise or Genus wise list of Lower Taxa and will drill down in both the cases. The drill down steps are given below:
   - For a given Family name, it will describe the family as given in Flora of India along with Genera links.
   - For a particular Genus, the information regarding Genus as per Flora of India along with Species links are obtained.
   - For a particular Species, the description, citation and distribution of the Species as per Flora of India will be displayed along with links to Varieties of the species, if any. It will also display floral illustration images, if any.
   - For a particular Variety of the Species, the description and distribution data as per Flora of India are displayed.
3) The drill down steps for browsing Genera will be exactly similar to Family drill down as described above.
4) The drill down process includes various taxon types, viz., Group, Family, Genus, Species, Variety, Sub Family, Tribe, Sub Genus, Sub Species, Forma, Section, Subsection and Sub Tribe, in floral hierarchical fashion.
5) Browsing Scientific Name will display the description and distribution of the Species as per Flora of India along with links to Varieties of the species, if any. It will also display illustration image, if any. Illustrations will be provided by BSI in JPEG format.
6) Advanced criteria specific search as well as full-text search of the volumes of book is also available.
7) Based on taxonomic data, knowledge based DST is used to make decisions.
8) In accordance with the results of searching the distribution of taxon is mapped to the geo-location thus getting the GIS based floral density distribution, distance, proximity, topography etc.
9) The resultant information is depicted essentially in the form of maps along with representation in textual or tabular form as an integral feature.
10) The architecture of the whole system is described in two aspects:
    - Black box architecture of the whole system.
    - Software architecture.

Fig. 2 depicts the Black Box architecture of the system.

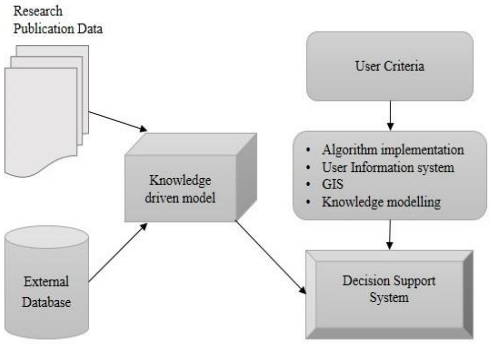


Figure 2. Black box architecture of the system

VI. System Architecture

To achieve this objectives the following step-wise interacting processes are required:

1) Collection of floristic data from all available sources such as publications, books and journals.
2) Collection of expert knowledge and information on their spatial distribution.
3) Development and evaluation of a decision support system prototype.

The software based solution is developed using web-based open-source technologies, viz., java, mysql, apache tomcat, apache lucene and struts 2 framework. Model-View-Controller (MVC) architecture is used is to separate the business logic and the application data from the presentation layer of the user. The main components of the MVC architecture are Model, View and Controller. Fig. 3 depicts the MVC architecture.

1) **Model:** It handles the data of an application. It also governs the business logic but remains unaware of the presentation layer.
2) **View:** It represents the presentation of the application. It is independent of the business logic and remains unaffected if the model changes.
3) **Controller:** Any request sent by the user passes through the controller. It is responsible for handling the request coming from view and passes it to the model for appropriate actions. After the action has been taken the controller is responsible for directing the appropriate view to the user.

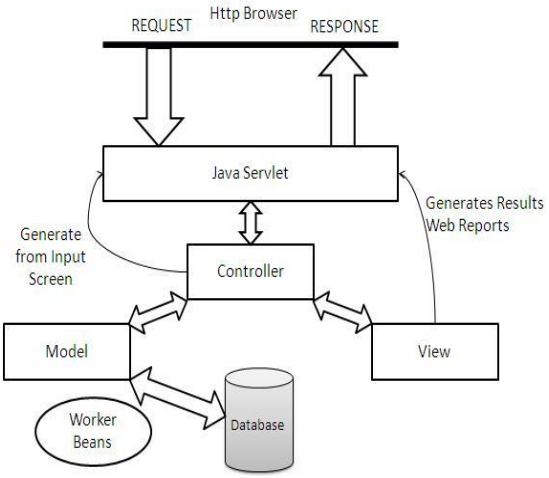


Figure 3. MVC Architecture

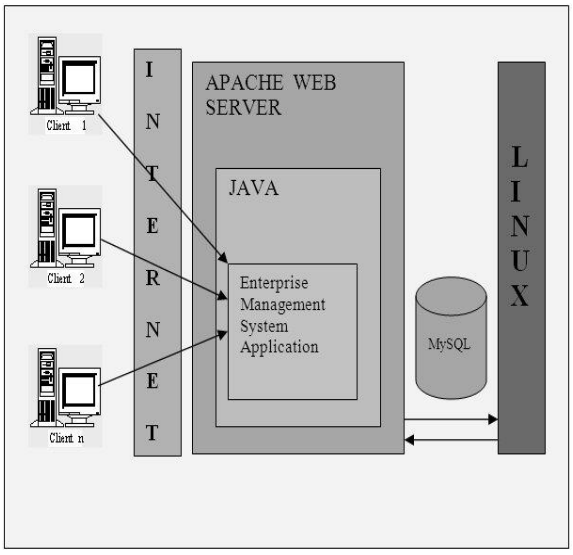


Figure 4. System architecture of the proposed model

The web based system architecture of the model in depicted in Fig. 4.

VII. Observations

An analysis have been made resulting in top ten dominant families of Kurung-kumey district of Arunachal Pradesh, India. Table I shows the same along with number of species and genera of the corresponding families. The density of Genera and Species of a particular family in the specified area can infer future decisions and predictions relating to various research oriented attributes of the family.

TABLE I. The Ten Dominant Families of Kurung-Kumey District of Arunachal Pradesh, India

Serial No.	Name of the family	Number of Species	Number of Genera
1	Poaceae	20	15
2	Asteraceae	19	17
3	Rubiacaee	18	12
4	Urticaceae	12	8
5	Fabaceae	10	5
6	Euphorbiaceae	8	6
7	Lauraceae	8	5
8	Zingiberaceae	7	3
9	Moraceae	7	1
10	Melastomataceae	7	4
	Total	**110**	**76**

VIII. Advantages and Disadvantages

The following are some of the advantages and disadvantages of the system:

A. Advantages

1) Research has demonstrated and substantiated reduced decision cycle time, increased scientists' productivity and timely information for decision making with the help of DST.
2) DST improves communication and collaboration among decision makers. Knowledge-driven DST provides a means for sharing facts and assumptions.
3) Research performance monitoring and ad hoc querying such systems can enhance research management benefitting financially.

B. Disadvantages

1) It reinforces the rational perspective and overemphasize decision processes.
2) Modelling of DST needs to make certain assumptions relevant to predefined situations which may be inappropriate. Better modelling and design of DST is the only way to avoid this potential problem.
3) DST decision lacks human intervention and innovation.

IX. Conclusion

The semantic network based knowledge modelling of floral diversity publication data has been found to be appropriate in converting unstructured data into structured format. The structured data has been used by the DST, which was developed using the emerging web-

based software technologies, to come up with research oriented analytical decisions. Relevant analytical data can be extrapolated by the use of statistical methods to predict future outcomes. Data mining concepts and algorithms, which refer to a class of analytical approaches, can be used to search for hidden patterns in a database to arrive at futuristic decisions. The model can also be used to design a potential tool for faunal diversity domain.

ACKNOWLEDGMENT

The authors are thankful to Col (Retd.). A. K. Nath, Executive Director of C-DAC, Kolkata for technical guidance, encouragement and providing required resources. The authors are thankful to the Director, Addl. Director, Botanical Survey of India and other eminent scientists for imparting domain knowledge and giving constructive suggestions during design and development of the DST.

REFERENCES

[1] S. Ravan, *Spatial Decision Support System for Biodiversity Conservation*, Team Coordinator, Geomatics Group, C-DAC, Pune, 2002.
[2] R. Mohemad, A. R. Hamdan, Z. A. Othman, and N. M. M. Noor, "Decision support systems (DSS) in construction tendering processes," *International Journal of Computer Science Issues*, vol. 7, no. 2, no. 1, pp. 35-45, March 2010.
[3] E. Turban and P. R. Watkins, "Integrating expert systems and decision support systems," *Management Information System Quarterly*, vol. 10, no. 2, pp. 121-136, 1986.
[4] C. W. Holsapple, "Decisions and knowledge," *Handbook on Decision Support Systems 1*, *International Handbook on Information Systems*, F. Burstein and C. W. Holsapple, Eds., Springer Berlin Heidelberg, 2008, pp. 21-53.
[5] M. Steyvers and J. B. Tenenbaum, "The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth," *Cognitive Science*, vol. 29, pp. 41-78, 2005.
[6] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Pearson Education, 2009, ch. 10, pp. 378-380.
[7] Karabatis *et al.*, "Using semantic networks and context in search for relevant software engineering artifacts," *Journal on Data Semantics*, vol. 14, pp. 74-104, 2009.
[8] E. Sasson, G. Ravid, and N. Pliskin, "Modeling technology assessment via knowledge maps," in *Proc. 47th Hawaii International Conference on System Science*, 2014, pp. 924-933.
[9] P. Skalle and A. Aamodt, "Knowledge-based decision support in well drilling," *Intelligent Information Processing II*, pp. 443-455, 2005.
[10] E. C. Way, *Knowledge Representation and Metaphor*, Kluwer Academic Publishers, 1991, ch. 4.

Sudhansu Sekhar Dash, (India, b.1969), obtained his M.Sc. (Botany, 1991), M.Phil (Community Ecology, 1994) and Ph.D (Taxonomy & Ecology, 1997) degree from Berhampur University, Odisha, India.
He has been actively contributing in the field of the Taxonomy of higher plants, Floristic and ecological studies in Eastern Ghats and Eastern Himalayan region of India for the past 22 years. He has authored 55 scientific research papers in Taxonomy, Ecology and Ethnobotany, co-authors 6 books, Edited three periodicals, and also has written more than 15 popular articles.
Dr. Dash is a life member in Indian Association of Angiosperm Taxonomy,(IAAT) Association of Plant Taxonomist (APT), East Himalayan Society for Spermatophyte Taxonomy and recipient of Dr. S.K. Jain Gold Medal, 2014, awarded by Society of Ethnobiology. Presently Dr is working as a Scientist in Botanical Survey of India and the Editor of the Journal Nelumbo (Official Journal of Botanical Survey of India).

Mina Desai, (India, b.1970), obtained her B.Tech. (Computer Science & Engineering) degree from Calcutta University, Kolkata, India.
She has been technically involved for the past 19 years with various computing projects in varied discipline contributing to Design, Development and Implementation of Database systems, Repository and Retrieval System, Web-based technologies solutions, Secure Software Development, Decision Support Systems using emerging Information and Communications Technologies. She has authored around 13 research papers in computing field.
Ms. Desai is a member of Computer Society of India. She is the recipient of "Young IT Professionals Award" conferred by Computer Society of India in the year 2000. Presently she is working as Principal engineer at Centre for Development of Advanced Computing, Kolkata, India.

Arghya Ghosh, (India, b.1985), obtained his M.Tech. (Computer Science and Engineering) degree form Kalyani Government Engineering College, West Bengal, India.
He has been actively involved for 6 years with various research projects in Soft Computing and technically involved in various computational projects contributing to Web-based technologies, Database Administration and Development and Repository-cum-Retrieval Systems. He has authored 5 research papers in Soft Computing and e-Business.
Mr. Ghosh is presently working as Project Engineer at Centre for Development of Advanced Computing, Kolkata, India.

Samaresh Das, (India, b.1976), obtained his B.Sc (Physics, 1997) from Utkal University, M.C.A (2003) degree from Biju Pattanaik University of Technology (erstwhile Utkal University), Odisha, India.
He joined CDAC, Kolkata (erstwhile ER&DCI, Calcutta) in 2004 and has been actively contributing to the projects related to Database and Data Warehousing technologies, Office Automation, dissemination of Agricultural, Livestock & Forestry information through PC & mobile devices.
Mr. Das has co-authored 1 conference paper on use of learning content Management System in the Domain of agriculture in vernacular. Presently he is working as an Engineer in CDAC, Kolkata and working on a project for development of repository of plant checklist in the Botany Domain.