Multidimensional Ontology modeling of Human Digital Ecosystems affected by Social Behavioural Data Patterns

Shastri L. Nimmagadda[1,2], Sashi K. Nimmagadda[1], and Heinz Dreher[1]

[1]Curtin Business School – Information Systems, Curtin University, Perth, Western Australia,
[2]Schlumberger, East Ahmadi, Kuwait.
e-mail: snimmagadda@slb.com; kiransashi@yahoo.com.au; h.dreher@curtin.edu.au

Abstract— **Relational and hierarchical data modeling studies are carried out, using simple and explicit *comparison* based ontology. The *comparison* is basically performed on relationally and hierarchically structured data entities/dimensions. This methodology is adopted to understand the human ecosystem that is affected by human behavioural and social disorder data patterns. For example, the *comparison* may be made among human systems, which could be between male and female, fat and slim, disabled and normal (physical impairment), again normal and abnormal (psychological), smokers and non-smokers and among different age group domains. There could be different hierarchies among which, different super-type dimensions are conceptualized into several sub-type dimensions and integrated them by connecting the inter-related several common data attributes. Domain ontologies are built based on the known-knowledge mining and thus unknown relationships are modeled that are affected by social behaviour data patterns. This study is useful in understanding human situations, behavioral patterns and social ecology that can facilitate health and medical practitioners, social workers and psychologists, while treating their patients and clients.**

Index Terms— **Ontology, digital ecosystems, social behavioural patterns, health and medicine.**

I. INTRODUCTION

Human ecosystems do not exist independently, but interact in a complex web of human and ecological relationships connecting all (human) ecosystems to make up the biosphere [1]. As represented in Fig. 1, it is a complex system of relationships, in which humans interact. These relationships exist in relational as well as nested hierarchies within several worldly contexts and domains. Humans individually or in aggregates behave differently in different contexts and domains based on language, culture, social, psychological and geographic locations in which these individuals or groups live. A household or university or an office may be examples of human systems in which several relationships are built with social systems to achieve certain social objectives or purposes.

Groups of humans, based on gender, body postures, skin colour, disability, marital status, occupation, poverty, habits and age are criteria and that can facilitate a more logical organization and physical representation of all relevant data in different conceptual and knowledge domains. An ontology is a specification of this conceptualization [2], describing such concepts and relationships that exist among humans. Groups or classifications can channel into a particular ecosystem based on the entities, dimensions and attributes.

II. PROBLEM STATEMENT

Data integration is a significant issue in the context of integrating wide variety and types of multidimensional data. Human and social ecosystems possess multidimensional data attributes along with their instances. Connecting these systems and extracting knowledge that can be interpreted by health, medical professionals and social workers is a real problem issue.

Families interact with the environment to form and make up an ecosystem. Families make good use of (as well as for society) biological sustenance, economic maintenance and balance the psycho-social functions. All individuals, families as a group, irrespective of their identity, are interdependent, especially on the usage of earth's resources. Consequently a balance between cooperation and integration of the ecosystem, satisfying the demands and needs of individuals and also respecting family institution as a social system/entity, their autonomy and freedom, is sought. Either as individual or group of individuals, human behaviour is connected to different dimensions of society. It is significant to understand and model these interdependent human situations from data patterns (Fig. 1).
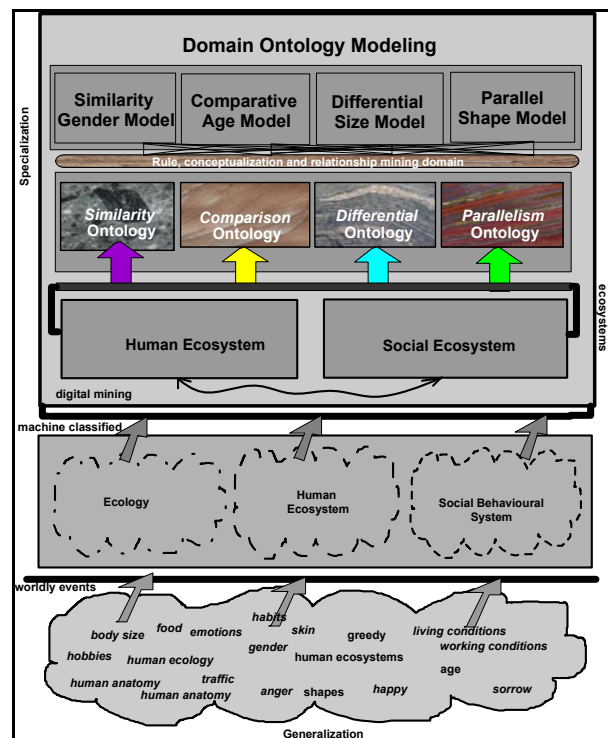

Fig. 1: Process model of domain otology

Human behavioural patterns are also unpredictable. In the ecologically integrated environment, these behavioural patterns (eco-pattern) have definite impacts on individual or among individuals. Humans express or react with a variety of emotions and feelings. Medical practitioners, social workers and psychologists, daily treat numerous people with abnormal behavioural patterns. These patients are entities, or described as conceptualization of several dimensions with several attributes, based on their sex, body postures, skin colour, habits (living styles, including food habits) and age. There may be volumes of data items and instances associated with these dimensions for several people and their associated behaviour or emotion attributes. *Behaviour*, *emotion*, *greed*, *anger*, *sorrow*, *happiness* dimensions which contribute to ecology, are interconnected to different dimensions and associated within human ecosystem. Each character or property which is scalable is considered as a dimension. All these dimensions are grouped into generalization and specialization categories. Manually handling these volumes of data and accessing them from different web and on-line/off-line sources are tedious processes. Authors propose ontology based data warehousing and mining to analyze these data for several categories and behavioural patterns, from several web sources more systematically in a way to ensure interoperability of conceptualizations. These human data patterns, human-ecological (interconnected) data patterns are integrated in a warehouse environment for data mining. Domain ontologies are designed for knowledge sharing and reuse among different contexts and domains, besides maintaining semantics.

III. PREVIOUS LITERATURE

Ecology associated with human emotional patterns has been well demonstrated in [1] narrating basic elements and causative factors of ecology. Management of ecology and issues surrounding human ecosystems have been discussed in [4]. Several domain ontologies relevant to medical and other engineering applications have been investigated by [5] and [8], demonstrating several multidimensional data structuring and data integration issues.

IV. METHODOLOGY AND FRAMEWORK

Data integration is a characteristic solution in digital system framework. Several attributes are conceptualized in these diversified multidimensional data structures of different domain ontologies. Digital ecosystem in the context of ecology, surrounding the human-being is essentially creating value by making connections between human-being and ecology of social behaviour domain. These domains could be generalization of *gender, human body posture, skin color, food habits* and *age*, and in the context of social ecology, any specific activity/action or business (different domains) within which human-being is connected. These are all embedded in a networked society which can be simulated in different IT collaborative frameworks. The relationships between human-being and ecology are conceptualized in multiple scalable dimensions; and each of which have definite interaction with other dimensions and attribution to different scales. In the context of multidimensional ecology, time and space are significant scalable dimensions.

Relational and hierarchical ontologies are proposed. As demonstrated in Fig. 1, several concepts are drawn from entities and dimensions based on *comparison, differential, similarity, parallelism* to develop relational and hierarchical ontologies. The *comparison* based approach is more specific and explicit, in which several properties of different data dimensions are compared. In other words, this approach compares data instances of attributes of multiple dimensions taken from several relational and hierarchical data structural models warehoused in an integrated environment (Fig. 2). In the case of *differential* concept, different instances (may be different in unique attributes) are drawn based on attribute/property strengths and sizes. The *similarity* concept is drawn from same data attribute property strengths and magnitudes. The attributes based on *parallelism* concept are drawn that are comparable to each other. Again, these concepts may have several permutations and combinations among themselves, based on logic, scale and context, based on which attributes and dimensions are conceptualized.
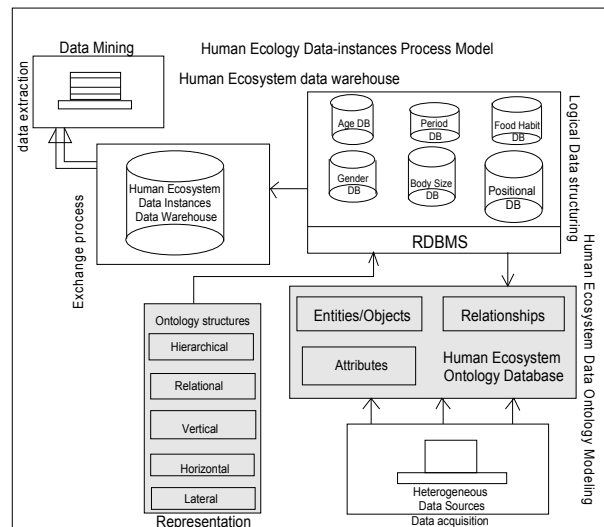

Fig. 2: An integrated ontology framework for building Human Ecosystem

A. Components, Objectives & Dimensions of Ecosystems

Adaptation is a continuing process in an integrated system (ecosystem) -each and every element contributing to the integrated ecosystem "*respond, change, develop*, and *act-on* and *modify* its environment." Measures and scales are significant indicators for social attainment and contentment. For example, there are socially defined and achievable goals which possess social values such as doctors treating patients, or need and desire fulfilment, rather than being outputs from the ecosystems. Social workers provide services to variety of their customers to full their needs that attain their survival goals. Entities that are inherited from an integrated and holistic approach preserve bio-diversity from genetic (generalization, in top/down hierarchy) to community level (specialization). These are open, dynamic and complex systems. The components of this system focus on dynamic interrelations, including social, biological, political, economical and physical features. Humans are parts of ecosystems with dynamic interrelations. Broad spatial and temporal dimensions are attributable to scaling these ecosystems. Several institutions, individuals and groups are parts of these ecosystems.

Male and female entities and their attributes can be 'in common' and 'uncommon'. Humans of 'skinny' and 'fatty' postures are other entities. Human emotions and feelings are behavioural attributes, having definite measure and scales. Documenting, acquiring these dimensions and attributes and logically organizing them are part of current scope of work. All these dimensions and attributes are hierarchical (Fig.3) and relational in nature.
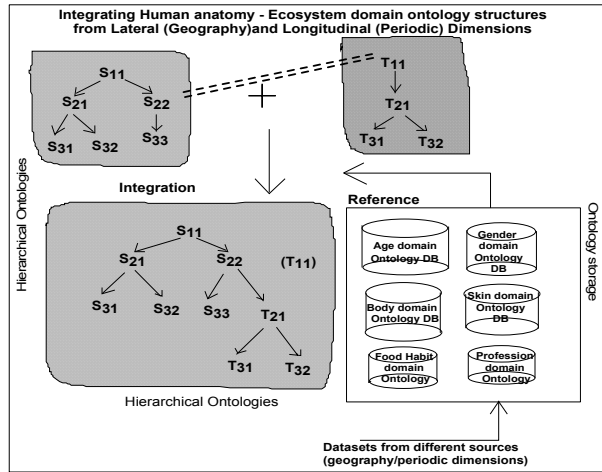

Fig. 3: Construction of hierarchical relationships from ontology storages

B. Multidimensional-base ontology modeling

Attributes are used under different contextual situations and semantics, such as *scalable, similar, equivalent, corresponding, analogous, matching*, and *comparable* to build relationships among data entities or dimensions of *sex, body postures, skin, habits* and *age* attributes of human ecosystems. For example, similar *sex* matching attributes modeled as *comparable* is a composite ontology in which conceptualized multiple dimensions and their associated attributes are inherited. *Short height* or *tall person* has scalable attribute too.
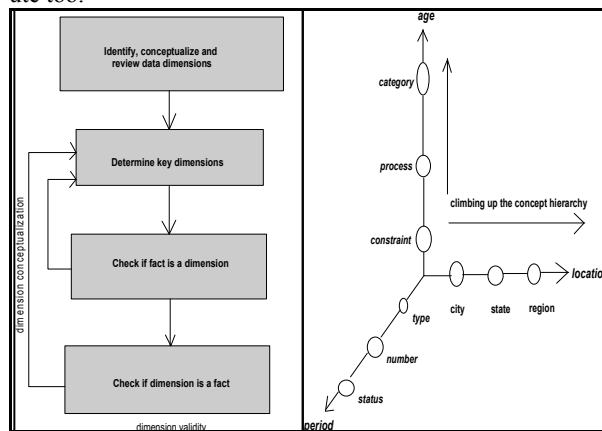

Fig. 4: Fact table identification process; multidimensional view of warehouse

"*Break in the chains of the infections*" and "*denormlization of relationships*" have similar analogies, may be in different domains, and specific data relationships are built within a multidimensional data model environment within the same analogy and context. This approach is more compatible and accommodative to a data warehousing approach. All the data dimensions that are deduced through this conceptualization are logical and validated within fine-grained data structuring approach as shown in Fig. 4.

In Fig. 5, the data connectivity between two different eco-systems is conceptualized thus establishing unknown relationships. As demonstrated in Figs. 6-7, star schemas narrate the age-behaviour eco-system models, based on deduced and validated multiple data dimensions (from Fig. 4) hierarchies. Key data attributes and their instances have been used to model human eco-systems with some instances being shown in Table 1.
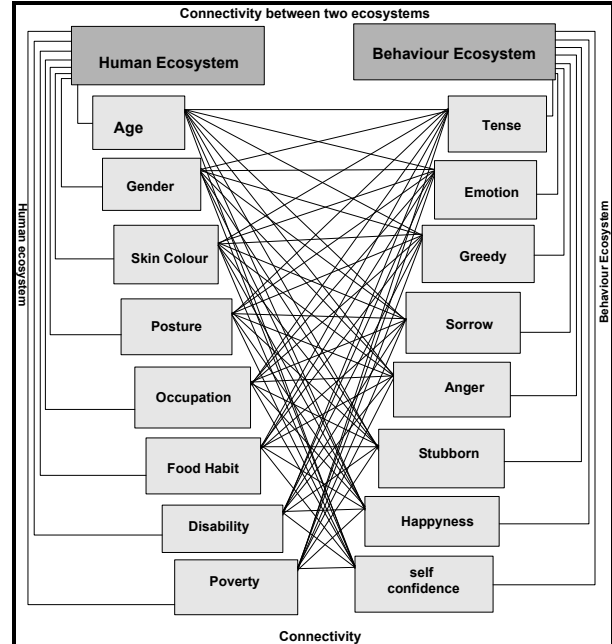

Fig. 5: Logically connected two ecosystems

Table1: Hierarchies for each dimension (with instances) of the current data warehouse

Period-dimension:
hour<day<month<year
location-dimension:
location-probe<city<state<region<country
age-dimension:
(infant, teen, adult, old) € all (age)
(13, 14, ...19) € teen
(45, ..., 50, 55) € old
(20, 21, 22, ..., 30) € adult
(1, 2, ...4, 5,...10) € infant
gender-dimension:
(male, female) € all (gender)
(10%, 20%, ... 35%, 40% ... 55%) € male
(50%, ..., 45%, 60%) € male
(35%, 23%, 45%, ... 35%) € male
(11%, 35%, ... 25%) € female

C. Data Mining and Knowledge Interpretation

Analogous to Global Positioning System (GPS) in global space domain, human anatomy is also described with several individual positioning sub-systems of human ecosystem domain, in which different dimensions are inherently positioned in human anatomy super-type knowledge domains. To this extent, more research work is in progress. However, the data views extracted from the human ecosystem data warehouse are interpreted for knowledge mapping in the following sections.
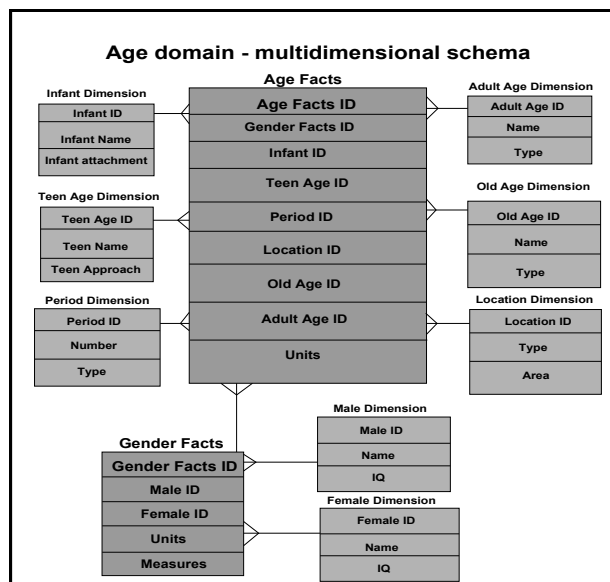
Fig. 6: Age domain Multidimensional Ontology connecting two different fact tables
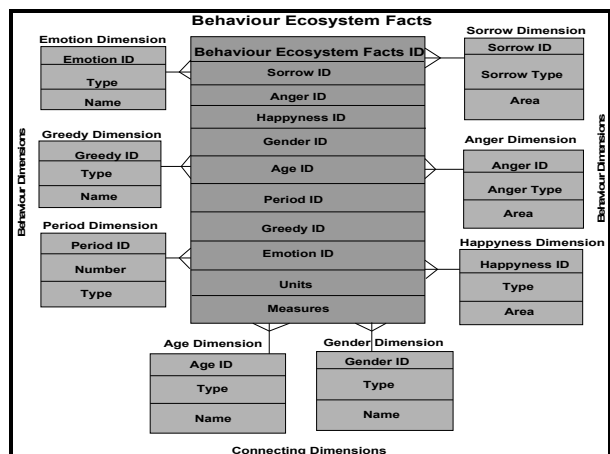

Fig. 7: Multidimensional star schema model narrating attribute dimensions

Factors affecting human ecosystem are scalable and measurable both qualitatively and quantitatively. These factors influence human anatomy [8] and the surrounding ecosystem, which is in turn interconnected through integrated domain ontologies. The knowledge mapped from these integrated frameworks is interpreted into measurable and scalable data views (Fig. 8). Semantic Web has standardized common domain ontologies that share and express worldly knowledge through *comparison, similarity, differential* and *parallelism* logics. In other words, knowledge building process may take advantage of domain ontologies and domain knowledge embedded within integrated human anatomy – ecosystem – ontology framework. Three main operations can be deduced [13], in data preparation stage, human anatomy ontologies are integrated with ecosystems by means of their heterogeneous data instances and appropriately guide the selection of data views to be mined. During data mining stage, domain knowledge is allowed with specification of constraints for directing the data mining procedures, for example narrowing of searching process. During interpretation stage (Fig. 8), domain experts of human anatomy and ecosystems validate and visualize the extracted data views.
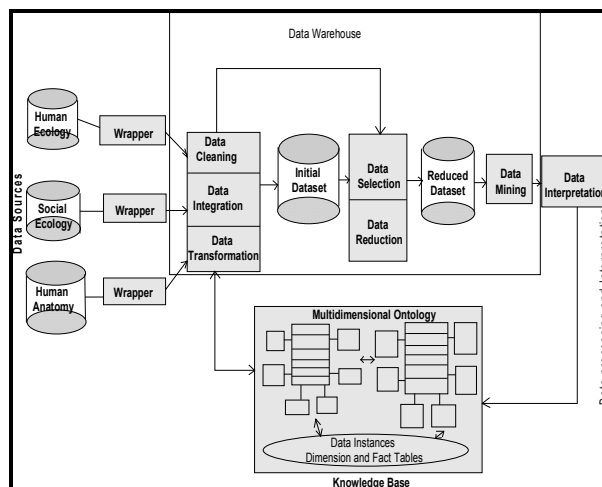

Fig. 8: Knowledge Discovery from Multidimensional Data warehouse

As a part of knowledge building process, it is important to understand the logically organized data dimensions. A constellation schema is prepared with human eco-system data instances as demonstrated in Fig. 9, interpreting multiple data instances.

Knowledge-base structure models: A *humanecosystem_socialecosystem* structure model Z for L (logic) is a 5-tuple (POSITION, AGE, GENDER, SKIN_COLOR, FOOD_HABIT, PERIOD). Social Ecosystem Model Structure Z1 for L1 (Logic1) is a 5-tuple (POSITION, EMOTIONAL, GREEDY, ANGER, SORROW, HAPPYNESS, PERIOD) Here S = U{(POSITION, AGE) R1}and U {(AGE, GENDER) R2} are domains of Z structure, and consists of the union of two mutually disjoint sets (POSITION, AGE), R1 and (AGE, GENDER), R2. (POSITION, AGE) is a set of individual entities of S and R is a set of relationships between (POSITION, AGE) and (AGE, GENDER) entities. R is partitioned in different ways as designer wanted it as R1 and R2, since the prior entity combinations are logically related.
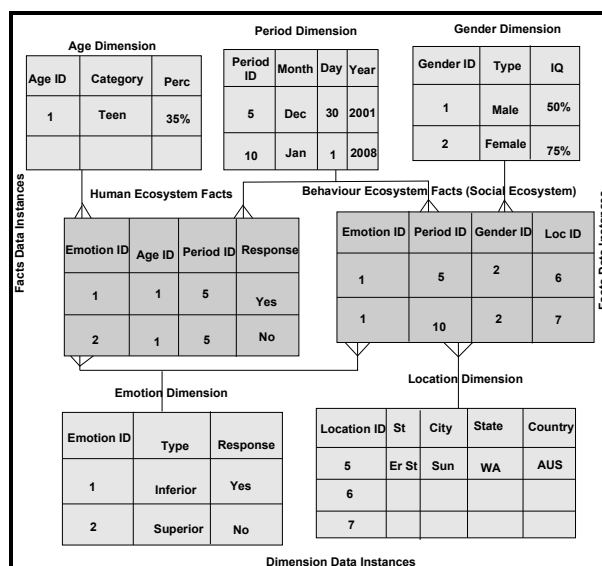

Fig. 9: A constellation schema connecting human and behaviour ecosystems

If this 2-tuple is to be interconnected, it could be done through relational or hierarchical relationships, such as:

S = U{(POSITION, AGE)R1} and U {(EMOTIONAL,

GREEDY)R2} are other domain of Z1 structure and consists of union of two mutually disjoint sets (POSITION, AGE), R1 and (EMOTIONAL, GREEDY), R2. (POSITION, AGE) is a set of individual entities of S and R is a set of relationships between (POSITION, AGE) and (EMOTIONAL, GREEDY) entities. R is partitioned in different ways as designer wanted it as R1 and R2, since the priority entity combinations are logically related.

Another structural model, *anatomyStructure_unitInfection* representing a region (A) composite (at generalization level) of all anatomy units is about locating an infection point (specialization level) in a anatomy system.

Anatomy=(A1, A2, A3... Ai); i=Anatomy unit numbers; A=Anatomy units
Am = U {(L1D, L2D...LmD) + (L1S, L2S...LmS)};
Lm.n=U {(Pm, (n, n+1, n+2) Pm, (n+3, n+4, n+5) Pm, (n+6, n+7, n+8) .)};
U= union; L = location; P = point; D = detection point; S = source point
m = number of treatment lines and n = number of infection/treatment points on each treatment line.

Data preparation – or pre-processing is aimed at quality controlling the data, data cleaning, transformation, reduction and data integration. Each has an impact on the other, when data inconsistencies are detected and rectified. Data integration is done using data warehousing approach, combined with ontology based multidimensional domain ontologies. Logically warehoused data are normalized (or denormalized for fine-grained structures), integrated and smoothed [9], which are ready for processing by data mining algorithms. Different data relationships are conceptualized [5], interpreted among several data attribute of human ecosystems. These ontologies are also responsible manipulating conceptualization and contextualization of data transformation processes. Ontology facilitates logical description of data, in away to reduce the number of dimensions without altering the integrity of original datasets.

Strategies that include reduction of data are **data cube aggregation**, **dimension reduction**, **data discretization**, and **data selection**.

Data cube aggregation produces data cubes for storing multidimensional aggregated data (e.g. extracted from a data warehouse) for OLAP (On-Line Analytical Processing) analysis [6]. For example, data on human relationships and socio-behavioural data held on millions of items, are aggregated into each specific domain ontologies.

A 3-dimensional data cube is created for a star-schema, which contains 12 x 8 x 7 = 672 cells. Stored within each cell is a behavioural data instance within a given category of human ecosystem for each month. If an average percentage emotional behaviour among an age group is to be computed, the cube contains a count representing these attributes. Some data views drawn from the cube are shown in Fig. 10. Each attribute of an OLAP cube may have one or more associated conceptualized hierarchies with multidimensional data structuring.

Dimension reduction leads to the encoding of data in a reduced format, with or without loss with respect to the original data set. For example, *similarity* dimension analysis is used for dimensionality reduction that applies to projections of initial data onto a space of similar dimensions. Again these similar dimensions are segregated as per their scale and magnitude.
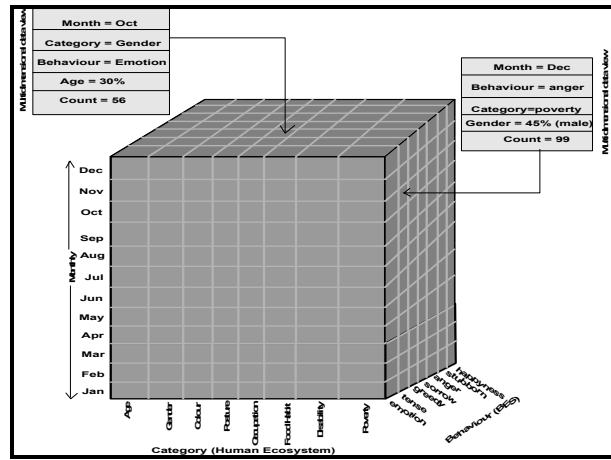

Fig. 10: Data views representation from multidimensional human and behavioral ecosystems

Data discretization is used to minimize number of dimensions (as per logical organization of their attributes, which can classify to a meaningful group) of each entity and consequently help interpretation of mining results. These classifications could as per scalable dimensions. In such cases, the range of attribute can be divided into several intervals (by means of histograms), which can further be iteratively aggregated into larger intervals. Scales are data dependent. If user has much understanding of the data attributes, an appropriate scale can be defined.

Data selection aims at identifying appropriate subsets among the initial set of attributes. This operation can be performed with the help of heuristic methods based on tests of significance or entropy-based attribute evaluation [13] measures such as the information gain. Data selection is one of the data reduction methods.

V. RESULTS AND DISCUSSION

In the digital world, human ecosystem is essentially about creating value by making connections between human and social ecosystems in different domains through support of different forms of collaborative IT frameworks such as data warehousing and data mining integrated approaches. Dimensions are subset, discretized and reduced to finer and specialized level.
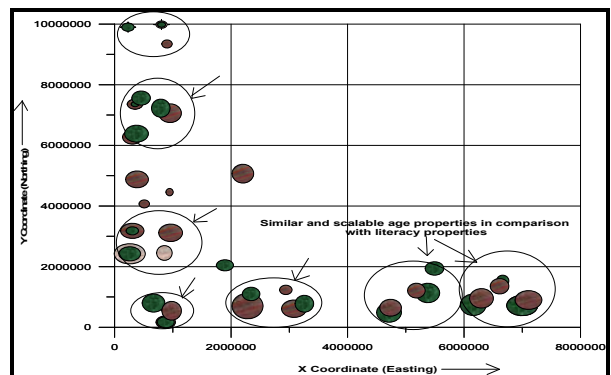

Fig. 11a: Data View showing similarity and scalable data properties in different regions (age vs. literacy)

The dimension data instances plotted in bubble plots have different scales, similarities and also dis-similarities. As demonstrated in Fig. 11 different attribute combinations have been plotted such as, *age* vs. *literacy*, *occupation* vs.

502

poverty and *poverty* vs. *emotion*. Different groups and sub-sets of dimensions have been identified with respect to geographic dimensions.

Similar, dis-similar and scalable bubbles represent several dimensional magnitudes and the aggregations of these groups of dimensional attributes.
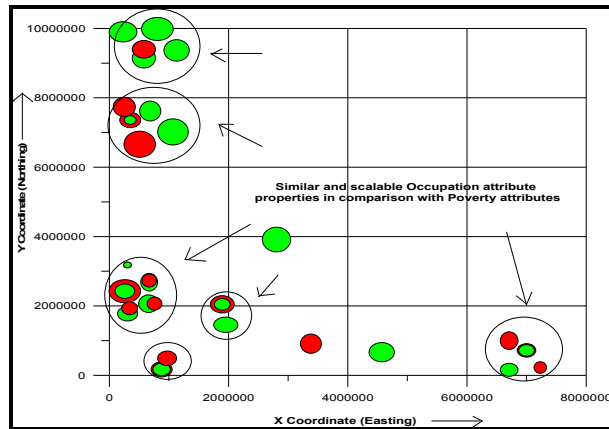

Fig. 11b: Data view showing similarity and scalable data properties in different regions (occupation vs. poverty)
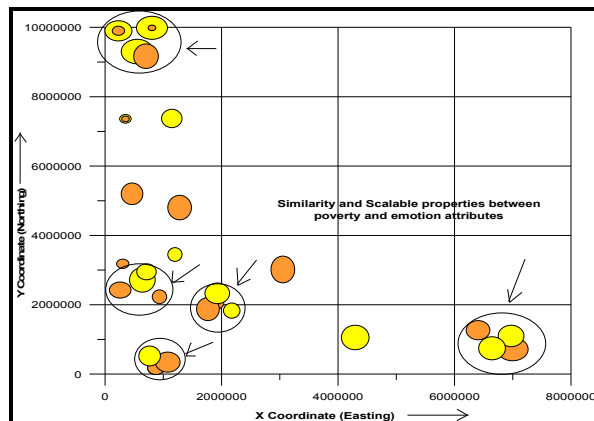

Fig. 11c: Data view showing similarity and scalable data properties in different regions (poverty vs. emotion)

In recent years, increasing use of high dimensional data that occupy large number of database tables with millions of rows/columns is getting popular. Also, large competitive demand for rapid build and deploy data-driven analysis is increasing with several data mining algorithms and solutions. A third trend is presenting the analysis of results to end-users in a form that can be easily interpreted to improve decision making. Ontology based data mining can emphasize scalable, reliable, fully automated and interpretable data structures that can address the data analysis challenges.

VI. Conclusions and Recommendations

Ontology based data warehousing appears to be highly effective, due to fine grained data structuring, effective data integration and interoperability of data dimensions in different application scenarios. Ontology handles conceptualization; contextualization and semantics among conceptualized data attribute relationships. Knowledge building from these approaches is effective and efficient. Data models described in the present case study are more flexible. Main

emphasis is model building will be addressing to development of data mining approaches, which are highly automated, scalable, and reliable. Analysis of domain ontologies and thus interpreting domain knowledge are more challenging issues.

VII. Scope and Future Work

Ontology models deduced in the paper have a scope of being extended to other areas of socio-economic systems, actually, any integrated system that is compatible to environment system, has a common base to many other ecosystems. The next article elaborates on some of the challenges that will need to be addressed to enable a whole new set of exemplary applications. Research is in progress, in designing and implementing the domain ontologies and their models in predicting climate changes and thus weather models.

VIII. References

[1] Anderson, E.N. [1996] Ecologies of the Heart: Emotion Belief and the Environment, New York and Oxford: Oxford University Press.

[2] Gruber, T. [2007] Collective Knowledge Systems: Where the Social Web meets the Semantic Web; Web Semantics: Science, Services and Agents on the World Wide Web (2007), doi:10.1016/j.websem.2007.11.011; http://tomgruber.org/

[3] Gruber, T.R. [1993] A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220; http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html

[4] Machiis, G.E. Force, J.E. and Dalton, S.E. [1994] Ecosystem management, Technical paper submitted to the Interior Columbia River Basin Project, University of Idaho, Moscow, Idaho 83843.

[5] Nimmagadda, S.L. and Dreher, H. Ontology-Base Data warehousing and Mining Approaches in Petroleum Industries: chapter XI in Negro, H.O., Cisaro, S.E.G., and Xodo, D.H, (Eds.), Data Mining with Ontologies: Implementation, Findings and Frameworks, pp 211-236. Information Science Reference, IGI Global, Hershey, PA, USA, 2007. http://www.igi-pub.com/reference/details.asp?ID=6844

[6] A.K, Pujari. *Data Mining Techniques,* University Press (India), 2002.

[7] Cadez, P. Smyth, and H. Mannila. Probabilistic Modeling of Transaction Data with Applications to Profiling, Visualization, and Prediction. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2001)*, pages 37–46.ACMPress, New York, NY, 2001.

[8] Nimmagadda, S.L and Nimmagadda, S. K, and Dreher, H. (2008) Ontology based data warehouse modeling and managing ecology of human body for disease and drug prescription management, a technical paper presented and published in an international conference of IEEE-DEST, held in Phitsanulok, Thailand, 2008.

[9] Hoffer, J.A, Presscot, M.B and McFadden, F.R (2005). *Modern Database Management*, Sixth Edition, Prentice Hall.

[10] Shawkat Ali, A. B. M. and Wasimi, S.A. (2007) Data mining: Methods and Techniques, Thomson, p. 299.

[11] Nimmagadda, S.L. and Dreher, H. and Rudra, A. (2005) Ontology of Western Australian petroleum exploration data for effective data warehouse design and data mining, a paper presented and published in the *proceedings of the 3rd international IEEE conference on Industrial Informatics*, held in Perth, Australia, August, 2005.

[12] Plastria, F. Bruyne, S. D. and Carrizosa, E. (2008). Dimensionality reduction for classification: Comparison of techniques and dimension choice, published in the 4th International Conference, *ADMA 2008*, Chengdu, China, October, 2008.

[13] Adrien Coulet**,** Malika Smaïl-Tabbone**,** Pascale Benlian**,** Amedeo Napoli and Marie-Dominique Devignes Ontology-guided data preparation for discovering genotype-phenotype relationships**,** *BMC Bioinformatics* 2008, 9(Suppl 4):S3.