International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 3, March - 2013

Managing Information Integrity In Information Systems

Reema Khurana
Professor (Information Systems Management)
Institute of Management Technology,
Ghaziabad, Uttar Pradesh, India

Purpose

The paper argues that, in order to ensure information integrity (I*I) - i.e., information with accuracy, consistency, reliability- is maintained, need, is to attach acquisition cycle and utilization cycle to the database. This converts the database of the traditional static environment to an information system coupled with an information base impacted by the system environmental factors, namely, complexity, change, communication, conversion, and corruption.

Design/Methodology/Approach

In order to achieve the purpose of ensuring Information Integrity in Information Systems the following research paper models acquisition cycle (and utilization cycle) as comprising multiple decision stages, which are then further developed to present the acquisition cycle as an information envelope with multiple information bases .The information acquisition cycle discussed in the paper models the information envelope (which is designed as the information identification, origination, and evaluation methodology) into an information system coupled with the information bases characterizing the transformation of information requirements mentioned above. It is shown that for this it is important to identify the implicit and explicit informational requirements in any evolving, conflicting and ever changing business environment. Information Topology plane has been introduced which will encompass all the stakeholders in the system and their characteristics. MERU approach has been suggested which combines the storage of information with a view to preserve integrity (all the concepts discussed above i.e. the acquisition cycle, utilization cycle, information topology plane, information envelope and the decision stages therein are part of the MERU approach).

Findings

The findings of the presented analytical research below are the model –MERU which is stated and can be used for ensuring Information Integrity of the Information Systems, also a systems view of Information Integrity has been presented.

Originality/value

The paper is original and adds value, as this research has not been found as published in any of the journals and has been developed by the authors.

Key Words: Changing environment, Complexity, Information systems modeling, Information bases, Consumer needs

1.0 Introduction

Database systems and information systems go hand in hand. The issues on account of system environmental factors of Complexity, Change, Communication, Conversion and Corruption (5Cs) affect the integrity of data being processed in the information system and being stored in the database [Mandke,Nayar,2000]. These result in information stored with users, which is with loss of Information Integrity (I*I). Information Integrity is defined as accuracy, consistency and reliability of information. The traditional databases have considered integrity but not with reference to the context and situation of database application, which in turn bring in the implication of environment. Integrity has largely been used as data integrity referring to fixed relationship, rules and domain [Elmasri, Navathe, 2004]. The traditional database defines integrity and also suggests ways to control the same; however it practices that the data and information once validated is good for future use i.e. no error is anticipated. This paper proposes a model in which information is continuously originated to reduce system uncertainty in the wake of changing environment. The information originated or produced is compared against a set of standard values for computing error. These error values are then used to calculate the value of integrity of information. The discrepancy or gap between calculated and desired integrity values of information then provides the basis for integrity control to decide on information origination for storage and retrieval.[Khurana, Mandke, 2009]

2.0 Information Integrity in Databases

The database forms the core of any information system. System environmental factors are outside the logical environment of the traditional database design view and, hence, errors (as a result of them) are not amenable to control by data integrity mechanisms normally considered at the DBMS design stage. Information Integrity (I*I) is dependability and trustworthiness of information and controlling it is a key factor for determining strategic business advantage. Its attributes are accuracy, consistency and reliability of IS and information there from [Mandke, Nayar, 99]. The following paper proposes to discuss the analytical aspects as under: requirement to model a system as a potential source of information, identify decision situations, identify a new information model along with acquisition and utilization cycle , uncertainties in data management leading to loss of Information Integrity, systems approach to error reductions, requirements to view information as a bundle of interrelated attributes, cost benefit analysis of achieving optimal Information Integrity and a method of quantitatively identifying information value.

3.0 Information Base Management System (IBMS)

For business drivers of globalization, international competition and continuously changing customer requirements that are increasingly becoming local and instant, the need is then for a database system that includes accuracy, consistency and reliability along with a reference to the context which emerges as the most important issue. Also it has been found that quality which is a one time paradigm is not enough as the employees treat quality norms only as an artifact and do not implement them , thus what is required is an ongoing phenomenon. (Iden, 2012)

The contention of database systems supporting open systems is not to approach openness from the platform point of view, but to indicate that the openness is to the changing context of the data. Since more and more data needs to be acquired, validated stored and utilized in the database. These processes need to be attached in the form of acquisition and utilization cycle along with the current databases; giving a flavor of Information System to our traditional database system. Thus the term Information Base Management System can be introduced [Khurana et al, 2007].

Acquisition cycle and utilization cycle attached to the database are depicted in fig 1.

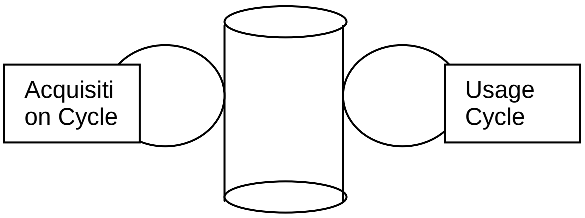

Fig 1: Traditional database with an acquisition and utilization cycle

It introduces the concept of error information base. The error creeps in the acquisition process when the information acquired is not as per the environment any more. This error information base then forms the basis of I*I analysis and subsequent integrity calculation in the information base.

The real world problems are open ended problems (Fabiana et al , 2012), and by no means, a complete problem can be solved. IS developed and deployed in a specific environment is always confronted with such open ended problems, as a result of the uncertainties in the environment. A blind eye to such problems results in significant drop in the effectiveness and efficiency of IS. The designer of an IS thus has to draw a boundary to focus on that particular problem (problem information), making some environmental factors internal while others as external to the system. The traditional design of IS is done in isolation, assuming the environment to be static, not considering the dynamism of reality. The traditional database assumes that information once stored is good for future use i.e. no error is anticipated. But for competitive advantage the IS design should consider the environment which is evolving, conflicting and ever-changing. This calls upon for the designer to anticipate various environments where the IS would be deployed in and also of the neighboring IS which have direct and indirect impact on that IS. The mechanism to study this has been suggested as the Information Topology Plane.

3.1 Information Topology Plane

Information Topology plane is an abstraction to account for the environment discussed above. It is developed through an iterative and continuous process (as the business is a continios process) (Kasim et al ,2012) of information origination, evaluation and processing. It is enriched with information envelope [Mandke, Nayar, Malik, 2001]. Information envelop suggests layers of information for each data element (data element is the smallest identifiable unit of information on which IS can be developed; it could be a single entity or an information network of entities and transactions like objects (hardware and concepts),people, software, rules, norms, policies, financial mechanisms legislations etc.) . The attributes of Information Topology have been shown in fig 2

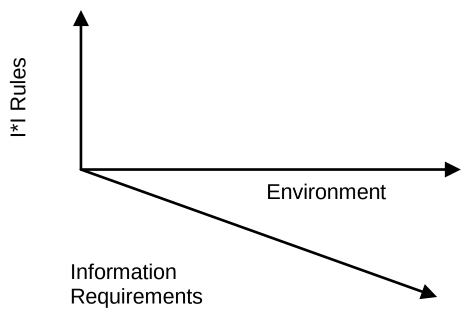

Fig 2: The attributes of Information Topology plane

Each layer has informational processes and decision mile-posts. This reduces the problem complexity while ensuring effectiveness and economic solutions. That is to say that how much of information should be originated in the topology. To this the answer is that corresponds to the information requirement of the dynamic decision stages. The varying types of requirements of an information system can be identified. These requirements are stated as problem information, problem environment information, and alternative information.

3.2 Requirements in Information Topology Plane

Broadly speaking the requirements can be listed as:

- Requirements with deficiencies.
- Shifting requirements i.e. shifting in conceptual objects which can be
  - Informational objects, mainly to be found in service sector
  - Functional objects – desired and maintainable.
  - Performance based objects.
- Requirements acceleration.
- Requirements coming with delays
- Transfer of requirements, horizontal, mobility of requirements
- Combination of requirements.
- Requirements with vertical mobility, telescoping(moving from lower level requirements to higher level requirements missing out on some in between subsequent levels)
- Decline in requirements.
- Evolving requirements – changing priorities.
- Conflicting requirements, complex requirements.

3.3 Decision Stages

This view of requirement is pregnant with design bases for error tolerant Information bases with Integrity .These requirements transform into the goal set at various decision stages. These goals sets then need to be stored in the information base (Yuelin, Nicholas, 2008) .Subsequently to meet each requirement (transformed as a goal set) data should be accumulated (originated and acquired). How far the requirement is met and how much more data is required in the design bases is then calculated, acquired and utilized. Process of transformation goes through various decision stages. The decisions stages can be identified as D0-D22. They are being enumerated as D0-D22 so that classification in to groups can be made:-

- D0 – Obtaining current basis data/information on requirements of: (a) Recipient (customer) under consideration (covers objects [concrete, abstract], humans, rules, norms, commands, etc.), (b) Business process costs and capabilities, (c) Questions, etc.
- D1 - Based on long-term goal set, determining Positive/Negative Goals, General/Specific Goals, Clear Goals, and Implicit Goals.
- D2- Transforming Negative Goals into Positive Goals.
- D3 - For Positive Goals identified, setting Intermediate Goals.
- D4- For the problem solving situation, identifying environmental anomalies or malfunctions that will emerge with delay.
- D5 - Given the malfunctions that will come with delay, determining what must remain unchanged, i.e., identifying environmental anomalies that must not occur in the process of problem solving.
- D6 - Based on (D5), delineating multiple goals to make implicit problem solving goals explicit.
- D7 - Based on Specific Goals (D1), (D3), (D5) and (D6), determining many factors and multiple criteria.
- D8 - Based on [(D3), (D5), (D6), and] (D7), determining independent goals.
- D9 - Based on (D8), deciding delegation (contracting), identifying uncertainties in delegated decision-making, and deciding operable goal statements.
- D10 - Based on [(D3), (D5), (D6), and] (D7), determining information about interdependent goals, which are positively linked.
- D11 - Based on (D10), selecting central goal from amongst positively linked goals and deciding operable goal statement.
- D12 - Based on (D10), deciding ranking of positively linked goals without time pressure and selecting operable goal statement.
- D13 - Based on (D10), deciding ranking of positively linked goals with time pressure and selecting operable goal statement.
- D14 - Based on (D3), (D5), (D6), and (D7), determining information about interdependent goals, which are negatively linked (i.e., conflicting goals).
- D15 - Based on (D14), choosing, from conflicting goals with uncertainty, the operable goal statement.
- D16 - From formally operable goal statement, defining planning & design constraints and opportunity spaces.
- D17 - From 'many factor' information variables characterizing problem complexity, culling out useful (relevant) information variables.
- D18 - Recognizing relationships (interdependencies) between culled out information variables

- D19 - Developing state transition model defining dynamic behavior of culled out state (information) variables.
- D20 - Within the framework of opportunity and constraints' spaces (D16) and based on the state transition model (D19), undertaking customized planning & design (i.e., unstructured and a periodic processing of factual information continuously obtained on current basis for the problem at hand for generating alternatives for evaluation)
- D21 - Evaluating alternatives generated at (D20) for their contributions to operable goals.
- D22 - Selecting flexible information decision for control implementation.

3.4 Environment

Further the environment would comprise of objects like (Giridhar Kumaran, James Allan, 2008)-

- People (like stakeholders, external user and internal user).
- Hardware.
- Software.
    - Requirements, analysis, design, development, technology (networks, web etc), testing, implementation, maintenance procedures.
- Norms.
- Policies.
- Rules.
- Procedures
- Concepts.
- Financial Mechanisms

These requirements, I*I rules and environment then form the attributes of Information Topology as shown in the figure above (Malcom et al, 2010)

As the result of considering the above stated mile-posts, we obtain a multi-layered Information Topology plane. The layers can be infinite in number but here we consider three for the understanding of the concept.

Layer 1: Problem Information.

Layer 2: Problem Environment Information.

Layer 3: Alternative Information.

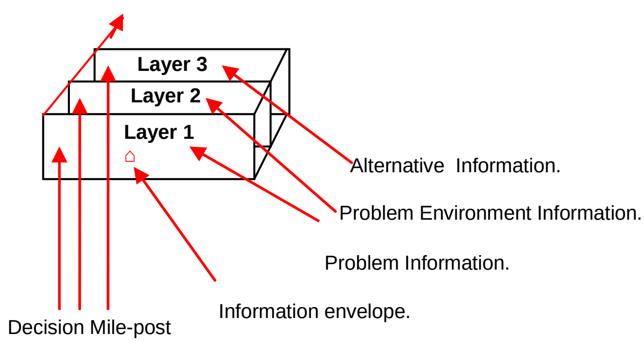

Fig 3: Architecture of IT.

Layer 1: Problem Information defines the problem. It clarifies or articulates the needs that must be met. The planning process of problem information has following four processes and their respective decision mile-post, namely,

- Analysis of needs.
  - D0.
- Determination of objectives.
  - D1-D7.
- Operational definition of objectives.
  - D8-D15.

Layer 2: Problem Environment Information reveals the opportunities and constraints that open or restrict the range of possible solutions and the general environment conditions, which will affect the contributions any potential solution will make to the various objectives. This layer consists of information about environment in which the problem is stated and would be deployed. The planning process of problem environment information has following two processes and their respective decision mile-post, namely,

- Basic studies: resources potential and needs.
  - D16-D17
- Analytical model.
  - D18.

Layer 3: Alternative Information is used for evaluating, selection of flexible information decision. The planning process for production of alternative information and decision mile-posts are:

- Systems synthesis.
  - D19-D22.

The layers are enriched with Information Envelope as shown in the fig 6 above. It is useful tool for information origination, evaluation and processing. Therefore, in developing problem information (Layer 1), first it is important to recognize that information is an envelope of values, objectives, goals and facts. Secondly, given long-term information, it is further required to obtain information in respect of operable value, objectives and goals' statement. In other words, one can originate information by defining long-term goal, many factors and multiple criteria, intermediate goals to account for problem complexity, independent goals, operable goals, interdependent goals, ranking and prioritization and conflicting goals.

4.0 IBMS –MERU Approach

Generating all the information on the above lines for the requirements, decision stages and the environment, now it is imperative to structure them in a database which can preserve integrity using the same (Sanda Harabagiu, Andrew Hickl, Finley Lacatusu, ,2007). The information base with integrity is being called Information Base Management System.

- The approach for designing the information base with integrity is called the MERU approach. The steps for creating and working with MERU information bases are:
  - Identifying the information topology plane.
  - Identifying the data elements in the plane, associating with each informational unit a trapezoid, each trapezoid encompasses decision stages (D0-D22) and information envelope, this information envelope suggests a methodology to originate and acquire information.
  - The information originated comprises of operable goals which transform as performance standards and their respective measurement criterion (performance standards are essentially measure of work performance and they are independent of personal attributes; in fact they are expectations from the data element).
  - These are stored as performance criterion in the MERU information bases. Any deviation of situational value recorded from the performance criterion is termed as error.
  - Errors in the IBMS form the basis of integrity analysis.
  - The Information base with Integrity (IBMS) uses the MERU approach as stated above.
  - MERU is an acronym for –
    - Main Information Base (M)
    - Error Information Base (E)
    - Rule Information Base (R)
    - Utilization Information Base (U).

These four information bases along with the information topology plane and the information trapezoid form the basis of the MERU approach for designing Information base with integrity.

Pictorially MERU can be depicted as in fig 4.

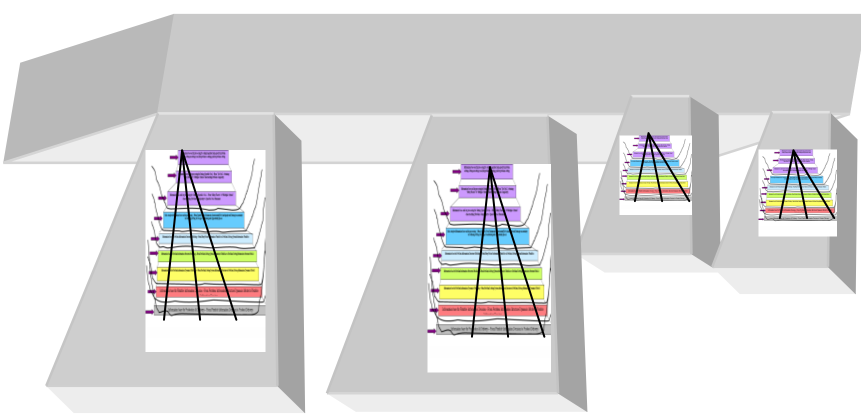

Fig 4: Depicting MERU on different contours of Information Topology plane

The box on top of fig 4 suggests the plane on which the informational requirements of the system rest. Each informational requirement is seen as a data element. These data elements could be rules, policies, people, commodities etc. The plane is representing the environment, which holds the contours of the data elements. These data elements are represented as trapezoids. What is the expanse of the plane or how many trapezoids it should encompass depends on the designer.

What are these trapezoids?

Each trapezoid is a suggestion of information origination, acquisition and utilization from the environment of an entity. On top only the long term goal of data element is stated. As we go down the layers in the trapezoid depict more and more information relating to each stage. This justifies the selection of trapezoids used for representing data elements .All this environmental information becomes part of the IBMS. These entities are all those which are participating in the information base.

Not only this information origination, acquisition and evaluation will occur in the acquisition cycle, quality assurance procedure but in the utilization cycle as well. The trapezoids depict the modularity of the flexible information base. This means that traditional database design involved only identifying the entity characteristics and then designing tables for them. MERU approach to design IBMS involves another structure before the tables are designed and data entered for each attribute. This structure states that all the tables in the information base will be defined in the MIB part of the information base. Not only this, each characteristic in the information base will consist of three additional attributes which will define their performance standards, performance criterion and situational value. As stated earlier any deviating situational value from the performance criterion will imply an error. This error needs to be recorded in the EIB. The integrity of MIB is calculated using EIB. Then the rules are applied on individual records of MIB, these rules are stored in RIB. After applying the rules the result is stored in UIB. The result may change only some of the attributes; the changed attributes are recorded in UIB. Then the integrity is calculated again, this time by picking up attributes from MIB and UIB together. Now the user may quit or in order to improve upon the integrity may originate more information. This methodology will provide the user with a number which is the information integrity indicator. In case the information integrity indicator is less than one the user may wish to originate more information. Also this entire process is happening at the design stage not at the implementation stage. MERU has been exemplified using case scenarios from healthcare, education, inventory and a problem of operator, supervisor manager. Each example has shown how origination, acquisition, evaluation and utilization of information with integrity leads to improved decision making.

The implementation can be done on any standard relational database platform..

The prime objective of MERU approach is to provide a mechanism to ensure integrity in database. Accuracy, Consistency and Reliability have been recognised as attribute of I*I.

MERU approach suggests the origination of maximal information for the purpose of preserving integrity. But a question which is always there on an analyst's mind is how much information to originate. Following cost benfit analysis gives an intrusion into this question.

5.0 Systems View of I*I

Further to the above discussion one can arrive at a systems view of Information Integrity. It shows an integration of all the concepts addressed earlier (Gaby et al , 2011). The systems view shows that how acquisition, utilization cycle discussed in are associated with a database can be shown from the designer's, process and user's view (Chikezie, Kiran Jude Fernandes,2012),. All the views will have Problem Information, Problem Environment Information and Alternatives Information and various decision mile posts associated with them. MERU can be associated with each data element and in the three different views mentioned.

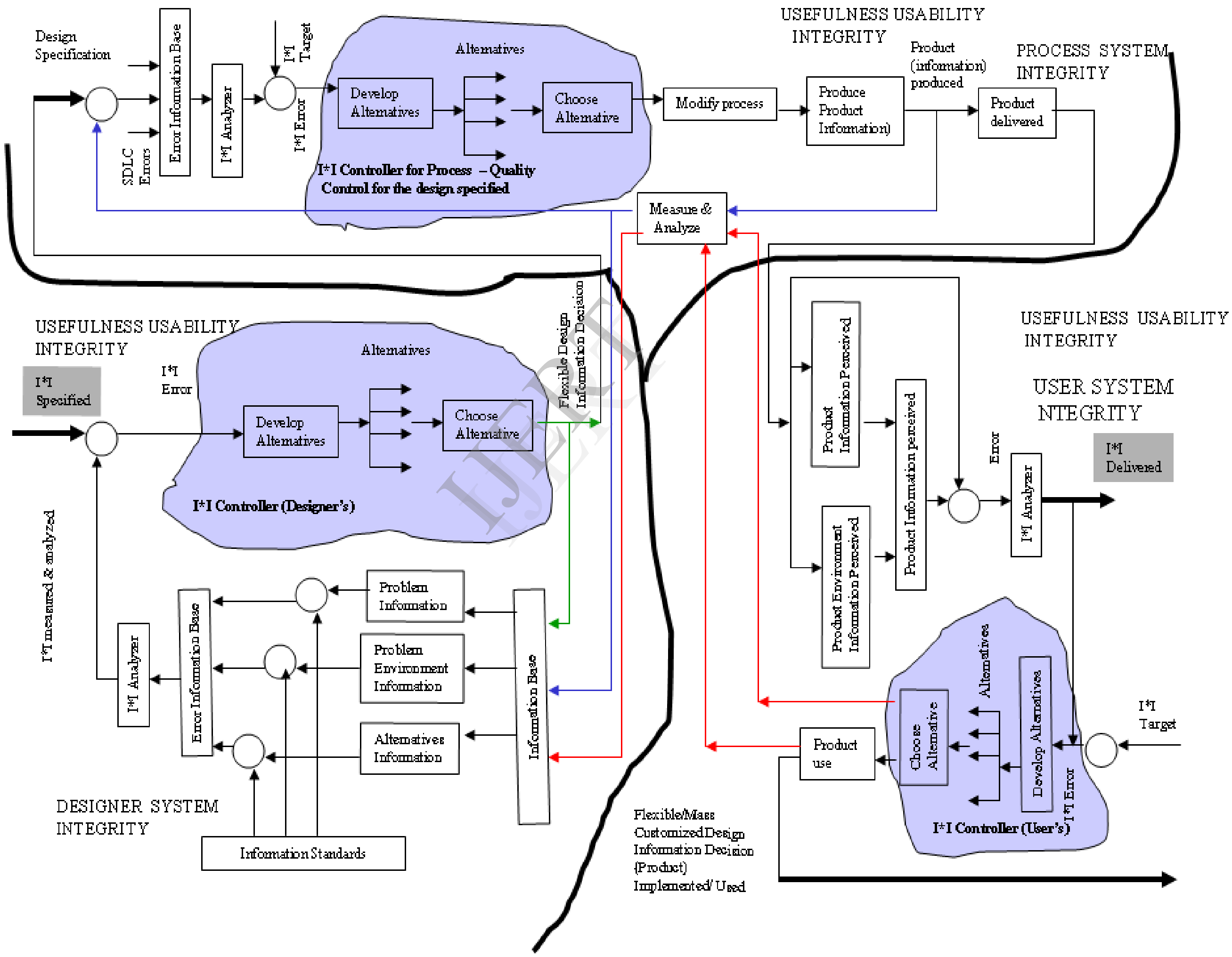

Fig 6: A systems view of Information Integrity

6.0 Conclusion and further research directions

This paper has identified the I*I issues in open information systems and the databases lying there in. It has then suggested an Information Base Management System using MERU approach. This approach proposes to attach an acquisition cycle and utilization cycle to the traditional database. Also the model rests on an Information Topology plane which has contours for each data element identified. Data elements are connected to each other and are associated with information envelop. The MERU bases exist across the envelope for each data element. Subsequently the cost benefit analysis of the model has been suggested and a systems view has been given at the end.
Research direction suggested are identification of intermediate decision stages, prioritization and ranking of operable goals, applications of information envelope to specific domains, inclusion of performance evidences in MERU ,application of cost benefit analysis to individual domains and reflection of cost benefit analysis in the overall systems view of I*I.

7.0 References

Chikezie Okike, Kiran Jude Fernandes,(2012), Impact of information use architecture on load and usability, International Journal of Information Processing and Managemnt, Elseiver, Volume 48, Issue 5 ,pp 995-1016.

Elmasri and Navathe, Fundamentals of Database Systems,(2004) 4rth ed., Pearson Education, Delhi.

Fabiana Jack Nogueira Santos, Claudia Cappelli, Flávia Maria Santoro, Julio Cesar Sampaio do Prado Leite, Thaís Vasconcelos Batista, (2012) "Aspect-oriented business process modeling: analyzing open issues", Business Process Management Journal, Vol. 18 Iss: 6, pp.964 – 991

Gaby Doebeli, Ron Fisher, Rod Gapp, Louis Sanzogni, (2011) "Using BPM governance to align systems and practice", Business Process Management Journal, Vol. 17 Iss: 2, pp.184 - 202

Giridhar Kumaran, James Allan, (2008), Adapting information retrieval systems to user queries, International Journal of Information Processing and Managemnt, Elseiver, Volume 44, Issue 6, pp 1838-1862.

Jon Iden, (2012) "Investigating process management in firms with quality systems: a multi-case study", Business Process Management Journal, Vol. 18 Iss: 1, pp.104 – 121

Kasim Randeree, Ashish Mahal, Anjli Narwani, (2012) "A business continuity management maturity model for the UAE banking sector", Business Process Management Journal, Vol. 18 Iss: 3, pp.472 – 492

Khurana R, Kapoor N, Ramesh S, Mandke V.V(2007) ,Designing an Information Base with Integrity-An analytical approach, International Conference on Information Systems Technology and Management '07 edited by Sahni, Malik and Agarwal ,New Delhi.

Malcolm Clark, Yunhyong Kim, Udo Kruschwitz, Dawei Song, Dyaa Albakour, Stephen Dignum, Ulises Cerviño Beresi, Maria Fasli, Anne De Roeck (2012), Automatically structuring domain knowledge from text, International Journal of Information Processing and Managemnt, Elseiver, Volume 48, Issue 3, pp 552-568.

Mandke V.V and Madhvan Nayar (1999), Modeling Information Flow for Integrity Analysis, Conference on Information Quality edited by Yahg W. Lee and Giri Kumar Tayi, MIT, Cambridge, Massachusetts, USA, pp. 38-57.

Mandke V. V. and Nayar M.K.,(2000), Information Integrity Imperative for Competitive Advantage in Business Environment characterized by Uncertainty, Information Technology for Business Management, 16th World Computer Congress 2000, Beijing, China pp. 115-125.

Mandke V.V, Madhavan Nayar and Kamna Malik (2001), Information envelope and its Information Integrity Implications, in Sixth International Conference on Information Quality Proceedings, Edited by Elizabeth M. Pierce and Raissa Katz-Hass, Cambridge, Massachusetts, USA.

Mandke V.V and Madhavan Nayar,(2002), Cost Benefit Analysis of Information Integrity, International Conference on Information Quality Proceedings, edited by Craig Fisher and Bruce Davidson, Cambridge, Massachusetts, USA.

Mandke V. V and Madhvan Nayar ,"Beyond Quality – The Information Integrity Imperative ",in 8th World Congress on Total quality Management, Dubai, 6-8 October, 2003.Also accepted for publication in International Journal on Total Quality Management, Taylor & Francis Ltd., Volume 15 No. 5&6, 2004.

Khurana R, Kapoor N, Ramesh S, Mandke V.V(2007) ,Designing an Information Base with Integrity-An analytical approach, International Conference on Information Systems Technology and Management '07 edited by Sahni, Malik and Agarwal ,held in March '07 in New Delhi.

Reema Khurana, Vijay V. Mandke, (2009) "Business process modeling with information integrity", Business Process Management Journal, Vol. 15 Iss: 4, pp.487 - 503

Sanda Harabagiu, Andrew Hickl, Finley Lacatusu, (2007), Satisfying information needs with multi-document summaries, Volume 43, Issue 6, International Journal of Information Processing and Managemnt, Elseiver, pp 1619-1642.

Yuelin Li, Nicholas J. Belkin, (2008), A faceted approach to conceptualizing tasks in information seeking, International Journal of Information Processing and Managemnt, Elseiver, Volume 44, Issue 6, pp 1822-1837.