DIAGNOSING THE VIABILITY OF FUNCTIONAL ANALYSIS OF A GENE SEQUENCE PROCESS IN MULTIPLE-RECURSIVE BIOINFORMATICS EXPERIMENTAL WORK SITUATIONS

Roliana Ibrahim

Department of Information System, Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia
roliana@utm.my

Abstract: The Functional Analysis of a Gene Sequence Process (FAoGS) is a process which requires molecular bench scientists to predict the function of a selected gene sequence. In order to predict a gene function, scientists conduct multiple and complex dry laboratory procedures. The focus of this paper is to explain the adaptation of two models; a Viable System Model (VSM) and Integrated Information Behaviour Model (IIBM) for diagnosing the viability of multiple bioinformatics work procedures embedded in recursive work situations. Undertaking experimental work procedures in this kind of environment require scientists to access, use and integrate variety of biological data from heterogeneous information sources. This paper explains how the use of VSM and IIBM support the modelling of dry bioinformatics work procedures in multiple-recursive embedded experimental work situations. This model allows systems analyst to identify the operational of autonomous work procedures at different recursion levels of work situations. The results of this analysis will allow systems analyst to understand and establish the existence of cohesive and adaptive mechanisms during FAoGS process in order to suggest appropriate design of Integrated Information Repository (IIR).

Keywords: Viable Systems, Information Behaviour, Bioinformatics, Gene Sequence, recursive model

1. INTRODUCTION

The aim of this paper is to present the Viable System Model (VSM) and Integrated Information Behaviour Model (IIBM) concepts, which is believe to be applicable for diagnosing the viability of a complex process. The focus process in this study is the Functional Analysis of a Gene Sequence (FAoGS). During FAoGS process, bench scientists normally perform bioinformatics analysis experimental works. Traditional wet laboratory approach requires scientists test any hypothesis in wet conditions and use wet materials. Unlike traditional approach, bioinformatics analysis consists of experimental work procedures which require scientists use various types of biological data and computer based analytical

tools available from their rich information space. This type of work procedures has becoming an integral approach which requires the molecular-bench scientists to adopt it in their scientific research work (Goble, 2001). Integrating bioinformatics analysis in a traditional wet experimental procedure can limit any time constraints facing by the molecular-bench scientists when preparing a gene or protein sample.

However, integrating diverse biological information from heterogeneous information sources is the main issue arise in conducting bioinformatics analysis. Molecular-bench scientists are facing with syntactic and semantic heterogeneities issues when accessing and using information from bioinformatics resources (Archard et al. 2001). Although there are technologies such as Web Services (Chelsom et al. 2002) and Semantic Web (Goble, 2001) that could provide solutions to the heterogeneity problems, the focus of this paper is not to discuss technological solutions to the issues. The focus of this paper is to discuss the concepts of VSM and IIBM for investigating the cohesive and adaptive capacities during bioinformatics analysis work activities. These two mechanisms are considered important since their existence in a process could determine the viability of the target process (Espejo, 2003). In this study, VSM and IIBM are considered as suitable methods to diagnose a variety of information-based activities during FAoGS process.

2. VIABLE SYSTEM MODEL (VSM)

The Viable System Model (VSM) concept was originally introduced by Beer (1979). The concept of VSM is applied for diagnosing complexities in the organization of interest as a living biological system. It is considered as a holistic approach which helps in understanding dynamic interactions among elements existed within an organization. An organization can be referred to as a viable system if they have the capacity to maintain a separate existence (Beer, 1985), evolve and adapt to changing environments. As a viable system, an organization should be able to maintain their balance when facing with unexpected and chaotic events.

VSM focuses on diagnosing organizational structure as one whole system and in a recursive manner in order to be able to establish the degree of cohesive and adaptive connectivity among the systems' parts. To diagnose the organizational structure of interest using VSM, the organization is divided into two levels; sub-systemic and meta-systemic. The reason for dividing organization into two levels is to assists in understanding the structural mechanisms exist within it. These mechanisms are the cohesive and adaptive capacities of an enterprise which aims at making the enterprise in a stable and balance state in any sorts of environment. These mechanisms are also essential for organizational design.

The sub-systemic level consists of operational, coordinator and controller units, meanwhile the meta-systemic level consists of the controller, intelligent and policy maker units. According to Espejo (2003), VSM provides a model which could assists system analysts to analyse the potential of the operational units at sub-systemic level by understanding how these units handle their problem situations autonomously, under consistent monitoring of a controller. On top of that, VSM support the diagnosis of an enterprise at meta-systemic level using the controller together with an intelligent unit to identify the existence of mechanisms that could create balance and adaptability. The way VSM support the diagnosis is by looking at these levels recursively. Each operational unit at each recursion level has its local or internal autonomy and responsibilities.

However, for an enterprise to be an effective one, it needs to produce not only local autonomy among these units but global cohesion. Using VSM, diagnosing the adaptability mechanism is focusing at the meta-systemic level, and for the cohesion mechanism, it is focusing at sub-systemic level. To summarize, VSM is a powerful tool or instrument that could support systems analyst and designer to identify whether an enterprise have the capacity to self-regulate or self-organize in a dynamic and changing environments. Establishing the ability of existing cohesive and adaptive mechanisms in an organization is considered useful for discovering potential solutions for intelligent organizational design. This is the reason why this study adopted VSM. It is considered as a method that assists the analysis of existing organizational structure in order to design an integrated information repository.

3. SYNERGY BETWEEN VIABLE SYSTEM MODEL AND CYBERNETICS

The VSM forms synergy between its concept and the concept of cybernetics. It incorporates the principles of cybernetics introduced by Ashby (1964), which focuses on the science of control and communication. Cybernetics concept deals primarily with the behaviour of a system rather than focusing on the objects. The main themes within cybernetics systems are coordination, regulation and control between the elements of these systems and their environment. Therefore, incorporating cybernetics concept in diagnosing organization is to take into account the communication or interactions that taken place within the organizational processes.

On this principle, VSM introduces five functions known as; *Operational, Coordination, Control, Intelligent* and *Policy*. These functions are diagnosed under System 1 to System 5 with additional of System 3* and System 3-4 Homeostat as stated below.

System 1 - Operational;

System 2 - Coordination;

System 3 – Control;

System 4 – Intelligent;

System 5 – Policy.

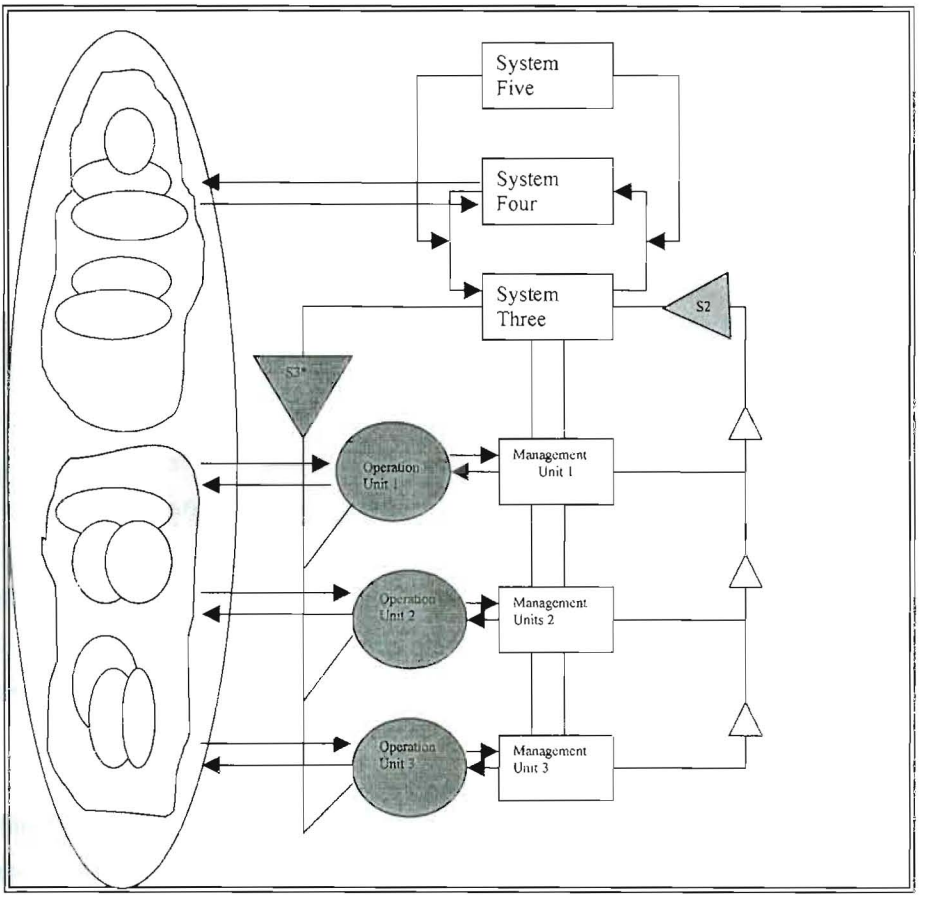

Figure 1 A Viable System Model

From Figure 1, the feedback loop between System 3 (The Controller Unit) and System 4 (The Intelligent Unit) demonstrates the existence of cybernetics principles in the model. System 4 interacts with its external environment and informs System 3 of the factors that could bring instability to the whole system. System 3 which receives input from System 4, controls the problematic situations and made appropriate decisions. System 3 informs System 4 of whatever decisions made and the feedback loop continues until the objectives set by System 5 are achieved.

4. INTEGRATED INFORMATION BEHAVIOUR MODEL (IIBM)

This model was introduced by Detlor (2003), which focuses on the *information use* environment. It focuses on general information use at an individual level together with their work setting. In this environment, the use of internet-based information systems is regarded as providing information for resolving individual problem situations. This phenomenon is similar to the work situation of the bench scientists. In general, bioinformatics analysis work involves the use of information from bioinformatics resources to solve scientific problems. Ideally in this kind of situation, bench scientists need to have easy access to the information they need and to be able to use this information for their problem solving.

In the context of this study, the work setting of bench scientists comprises the bioinformatics' routine procedures for their scientific analytical work. Those routine procedures are considered as critical components which influence the information need-seeking-use of the workers (Detlor 2003). This concept uses four variables that will assists in investigating the information behaviour of bench scientists while undertaking bioinformatics analysis activities within each experimental procedure. These variables are problem situations, dimensions, information traits and information use. In the context of investigating the information behaviour during the FAoGS process, the problems, situations and dimensions variables are referred to as the experimental work situations of the bench scientists.

The model shows that there are two main entities within the *information use* environment. These entities are the users and the Internet-based Information System. There are three stages in this model. Figure 2 illustrates the *information use* environment and the stages of information behaviour activities.

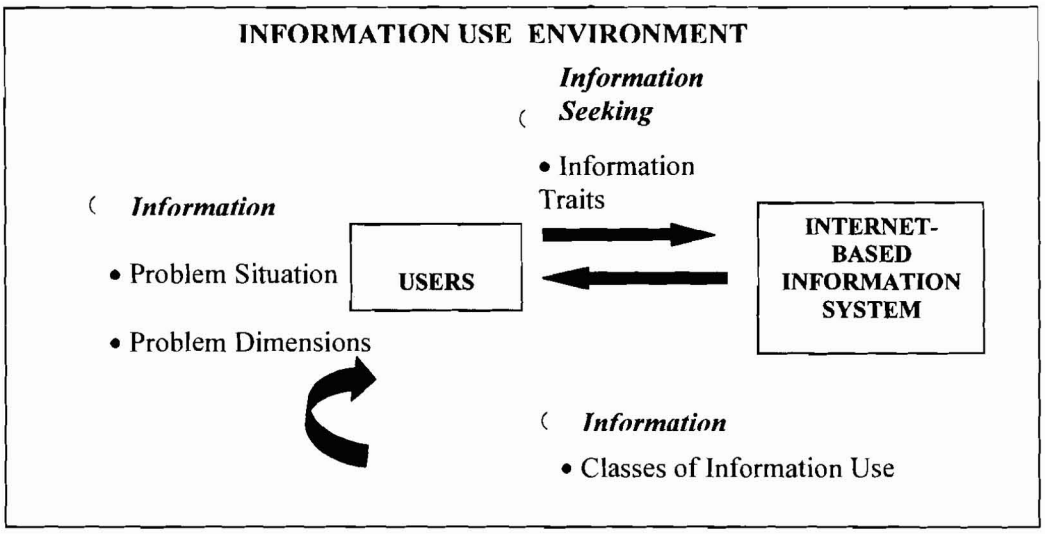

Figure 2: Integrated Information Behaviour Model (IIBM)

Stages in IIBM – IIBM proposes three stages in the study of information behaviour activities, which focus on two main entities, as mentioned previously.

First stage – information need: This is the stage where users face discrete problem situations within their individual work setting. Each problem situation possesses specific characteristics described as *problem dimensions*. By comprehending situations in terms of problem dimension, users would be able to establish the nature and types of information that they could elicit in order to solve current problems. When conducting bioinformatics analysis, the bench scientists are in the situation where they need to search for biological information from a host of heterogeneous bioinformatics resources. The information obtained not only supports bench scientists in their problem solving but also assists them in planning, designing and directing the subsequent experimental procedures.

Second stage – information seeking. In this stage, users use Internet-based information systems to seek relevant information for resolving current problem situations. In particular, they examine the actual information output provided by the Internet-based interface by establishing the value of such output against a 'checklist' of 'information traits'. Detlor (2003) describes information traits as characteristics of the information delivered to the users. Examples of these information characteristics could be *precise*, *hard* or *soft*, *current*, *single or multiple solutioned* and *diagnostic* in asking *what* or *why*.

The delivery of information to the user often relates to the information needs of that activity. Information alone has no purpose on its own and must be associated with an activity (Malmsjo 1997). Therefore, the role or objective of an information system is to deliver information related to a particular activity. It is important to consider the type of information utilised by the activity in question. Classification of the type of information used varies according to the users' objectives in their work activity and the information it requires. In a cybernetic system, it is essential for users to obtain an appropriate correspondence between actual information output and the information originally desired. This identifies with the idea that a cybernetic system would always aim for a state of equilibrium between information output and desired information. As such, the user's ability in determining the information traits or characteristics of the information output provided by the Internet-based interface would then be crucial for a cybernetic system to achieve this equilibrium state.

Third stage – information use: According to Detlor (2003), the *information use* stage starts when user has completed a scan of the information sources available from the Internet-based information system. The user selects and processes the information according to various classes of information use. Examples of these classes of information use are conformational, enlightening, factual, instrumental, motivational, personal and political, problem understanding and projective.

In the context of the work setting, there are three dimensions in which the use of information could be investigated (Choo et al. 2000). These dimensions are:

Affective: at this level, users are very selective in using information. Users will avoid using information that gives negative implications to themselves or others.

Cognitive: at this level, users select, use and process information based on their individual cognitive styles and preferences.

Situational: at this level, users use information based on the rules and routines that structured the tasks.

This study focuses on investigating information use based on the routines and procedures within bioinformatics work. As such, the attention of the investigation is on the situational dimension of information use within bioinformatics work settings.

5. BIOINFORMATICS ANALYSIS

Traditionally, molecular-bench scientists who conduct scientific analysis process starts their experimental work by generating their hypothesis, define this hypothesis in the format of analytical enquires and assess sample requirements for further sample preparation (Saunders, 1999). The following step is to further testing the prepared sample using appropriate techniques and methodology selected from either online publications or standard laboratory procedures.

Bioinformatics analysis is the term use to represent the conduct of laboratory or experimental procedures which requires scientists use variety and diverse types of biological data from heterogeneous information resources. The purpose of this procedure is to prepare sample of gene or protein, manipulating those sample data and further generating scientific hypothesis (Toldo and Rippmann, 2005). As such, the introduction of bioinformatics analysis procedures in scientific experimental work has changed the paradigm of undertaking scientific analysis process. The reason is because scientists has to make use existing biological resources which stores three categorical types of information; literature, factual and knowledge base (Kanehisa, 2001). In this new paradigm, scientists integrate bioinformatics work as supplementary method to traditional wet laboratory environment. The reason is, the use of available biological data from their information space will limit the time constraints in preparing the sample before further testing to the sample is undertaken in a wet laboratory. Examples of biological data and annotated information used by the molecular-bench scientists are genome sequences, DNA sequences, amino acid sequences, expression libraries, motif libraries, protein profiles and many others.

However, despite improving scientists' time of in their experimental work, there are complexities facing those scientists whenever they access and use biological information. These complexities are now discussed in the scenarios of molecular-bench scientists conducting bioinformatics analysis in Functional Analysis of Gene Sequence (FAoGS) process.

6. EXPERIMENTAL PROCEDURES FOR FAoGS

Functional Analysis of a Gene Sequence (FAoGS) is a process which aim at determining or predict the function of a protein in a single molecule. In this process, a molecular-bench

scientist who is interested in developing a new antibiotic against specific bacteria, extracts a gene sample from selected bacteria. The purpose of this task is to select and identify a gene which acts as a target for formulating a new vaccine. To be able to perform this task, a molecular-bench scientist needs to conduct and manage multiple bioinformatics work procedures in order to determine the characteristics of the selected gene in order to predict its function. The implementation of these procedures was initiated by the scientific analytical context and the work situations facing by a molecular-bench scientist. Figure 3 illustrates the work situations identified during the FAoGS process using IIBM.

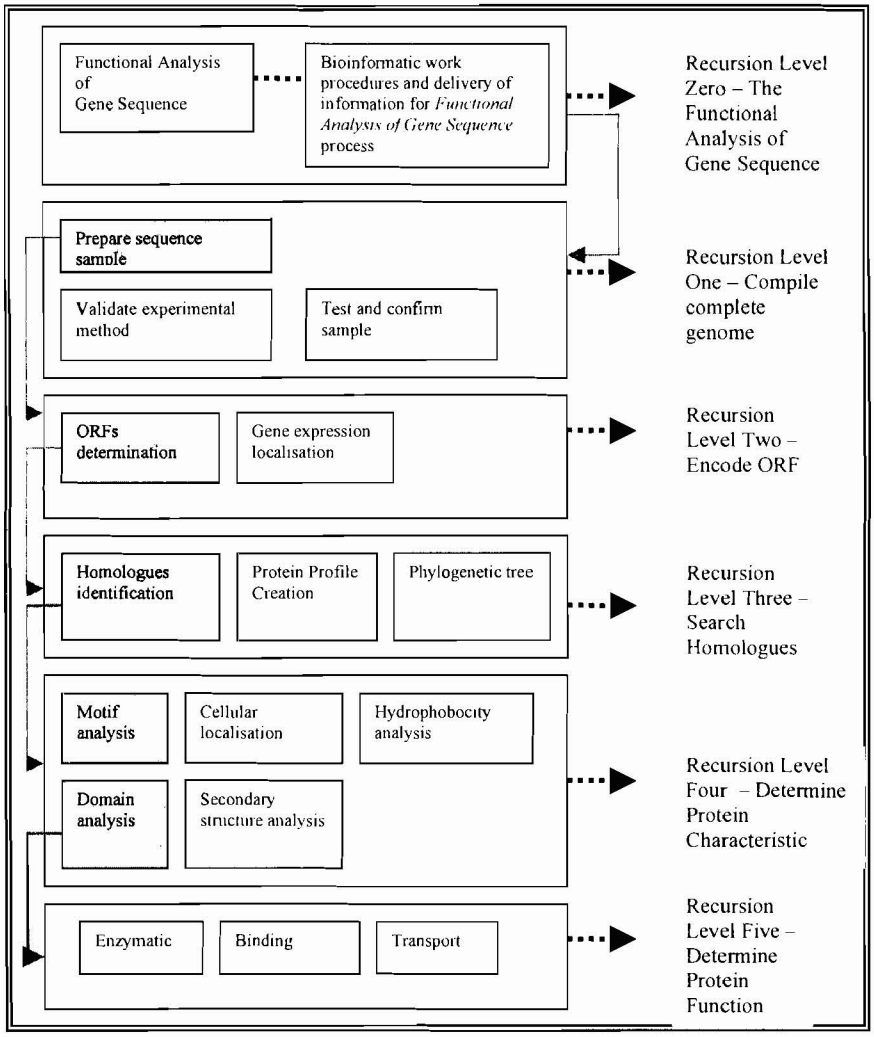

Figure 3: Process Recursion Model of FAoGS process

In each of the work situations as represented in Figure 3, there exists a variety of procedural components and interrelated information-based activities. These activities are conducted by the bench scientists in order to achieve the purpose of each experimental procedure. This

requires a bench scientist interact with multiple embedded information environments recursively. This shows that when undertaking FAoGS, scientists not only conduct a variety of experimental work procedures, but they too need to interact with a variety of different types of sequence data (eg. genome sequences, gene sequences, amino acid sequence, signal peptide sequences etc) and different types of computational tools and algorithms (eg. BLAST, FASTA, ANN, HMM, etc) within their information space. Therefore, the selection of VSM and IIBM in this study was to diagnose information used by the bench scientists in the organization of FAoGS process. Details analysis to Process Recursion Model (PRM) as in Figure 3 produced a basic process structure with a summary of diagnosis at each recursion level. Results of diagnosis are further elaborated in Section 9.

7. RELATED STUDY

Related study could be categorized in two different perspectives. The first perspective is the investigation of the experimental procedures in FAoGS process. The focus was on the bioinformatics analysis work. Investigation to the bioinformatics work was made in the context of bench scientists' information needs and use. In this perspective of the study, the most related work was done by Bartlett and Toms (2005). They investigated bioinformatics analysis procedures by adopting an integrated information behaviour and task analysis approach. Bartlett and Toms (2005) modeled bioinformatics analysis procedures in the form of three pathways of task hierarchy instead of recursively.

The second perspective of the related work is on the adaptation of VSM for diagnosing a complex situation. One related work found is (Ahmad and Yusoff, 2006). They adopted VSM to diagnose complex enterprise situation for Strategic Information Systems Planning (SISP). The aim of their diagnosis was to gather information requirements for SISP.

8. THE PROCESS RECURSION MODEL

The Process Recursion Model (PRM) as in Figure 3 is the primary representation produced before further diagnosis to each level of work situation was made. The PRM shows there are five multiple-recursive work situations in the FAoGS process presented in as Recursion Level 0. Further breakdown of this level such as Recursion Level One, Recursion Level Two and so on shows the work situations embedded in the previous recursion level respectively. As

mentioned earlier, this model was outlined by identifying the transformation of work situations during information-based activities in FAoGS process.

Interview data with ten molecular bench scientists from three departments in the Faculty of Science, University of Sheffield, UK was the main input for the development of PRM. The formation of PRM also takes into consideration the thermodynamics principle of central dogma organisation in a single molecule. This principle stresses on the genetic information flows through a single molecule which could be viewed from several levels of abstractions (Kanehisa, 2001). These levels of abstractions are represented in Figure 4. This diagram represents information flows at each of the abstraction levels in bioinformatics analysis work and information situations.

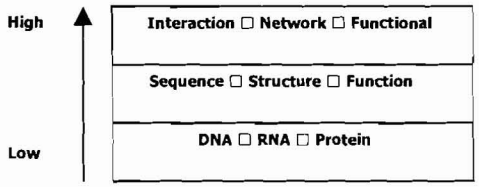

Figure 4 : The three levels of abstraction

The PRM derived from this study was represented at the low level of this abstraction which is focusing at dry experimental work procedures which transform DNA → RNA → Protein. The PRM shows that in FAoGS process, bench scientists first need to compile complete genome sequence. At this particular work situation (in Figure 3 it is noted as Recursion Level 1), there are three procedural units identified. These experimental procedures are preparing sequence sample, validate experimental method and test and confirm sample. Bench scientists continue perform these procedures until they achieve the right terminating conditions which transform them into another work situation.

9. RESULTS

The basic structure of the FAoGS process at each level of recursion was diagnosed for viability using VSM five prescribed function as mentioned earlier in the previous section. Figure 5 gives an example of the basic structure of Process at Recursion Level 0. In this figure, System Two (S2) is the function that coordinates the competition for data and resources among the operational units (System1).

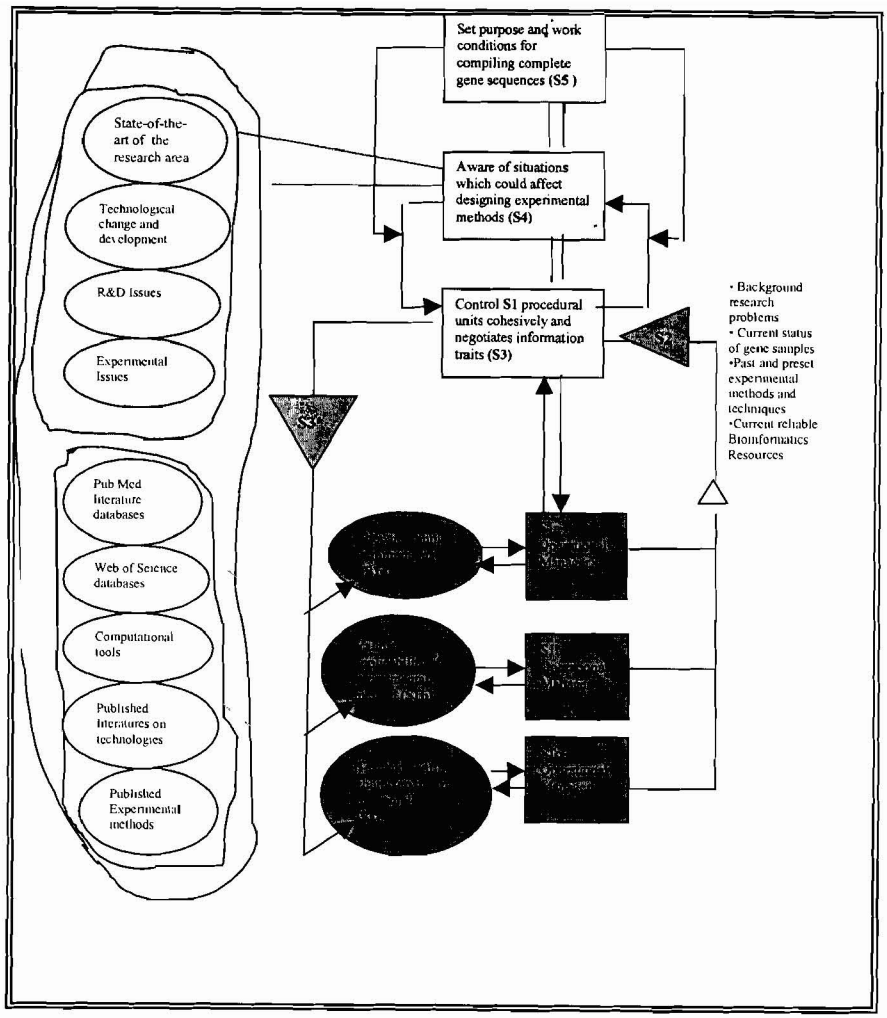

Figure 5 : Basic structure of Process Level at Recursion Level 0

9.1 Results of Diagnosis at Sub-systemic level

Results of diagnosis at sub-systemic level (System 1, System 2 and System 3) revealed coordinated information collected from diverse resources during each experimental procedure. In principle, each of this procedure requires different sets of data for determining the putative function of a gene sequence.

Result of diagnosis (for Recursion Level 0 to 2) shows the tasks and information coordinated by S2 at different recursion level.

Table 1: S2 at RL0 - The Functional Analysis of Gene Sequence

▪ search, track and collect past and present published literatures relating to the background problems and current status of the target gene
▪ search, track and collect past and present literatures relating to the technological development of bioinformatics resources
▪ search, track and identify possible resources to support current and subsequent work procedures

Table 2: S2 at RL1- Compile Complete Genome

▪ continuous tracking and retrieving of whole DNA sequence
▪ continuous tracking and searching for restriction sites.
▪ continuous tracking and searching of genome sequence
▪ continuous tracking and searching of primer parameters
▪ continuous tracking and searching for supporting literatures

Table 3: S2 at RL2- Encode ORFs

▪ continuous tracking and searching for matching gene and genome sequences
▪ continuous tracking and searching for gene loci positions
▪ continuous tracking and searching for ORFs positions
▪ continuous tracking and searching for ORFs protein coding region
▪ continuous tracking and searching supporting literature information

The tables show the tasks and information coordinated from various resources such as TIGR, NCBI and PubMed at three different recursion levels.

9.2 Results of Diagnosis at Meta-systemic level

Results of diagnosis at meta-systemic level revealed existing feedback loop between System 3 and System 4 and to establish whether these units are able to consistently keep the FAoGS process in a balance state or otherwise. The tables below show the result of diagnosis for Recursion Level 0 to 2.

Table 4 : S4 at RL0 - The Functional Analysis of Gene Sequence

▪ aware that using inappropriate experimental methods could introduce problematic situation. But, S4 incapable to predict any problematic experimental methods
▪ aware of using unreliable resources could introduce problematic work situation. But, S4 incapable to predict the use of resources that could create problematic work situation

Table 5 : S4 at RL1 – Compile Complete Genome

▪ aware of the situation which could affect retrieval of whole DNA sequence
▪ aware of the situation which could affect finding accurate restriction sites
▪ aware of the situation which could affect finding matching sequence
▪ aware of the situation which could affect finding accurate primers parameters
▪ incapable of predicting any of the above-mentioned events in advanced

Table 6 : S4 at RL2 – Encode ORFs

▪ aware with situation which could affect finding matching gene sequence
▪ aware with situation which could affect finding significant gene locus
▪ aware with situation which could obtaining significant ORFs
▪ aware with situation which could affect ORF sequence translation into protein sequence
▪ aware with situation which could affect finding appropriate literature information.
▪ incapable in predicting any of the above problematic situations.

The diagnosis reveals that System 4 (S4) function is not functioning effectively at most recursion levels since it is unable to predict any factors that could cause problematic situations to the bench scientists. This shows that currently S4 function which is supposed to be an intelligent function is incapable to predict any shortcoming events that could bring instability and failure to the process.

With limitation of adaptive mechanism during FAoGS process, there are tendencies that the FAoGS process is unable to evolve and adapt to any future or unpredictable changing environment.

10. CONCLUSION

This paper has outlined and explained the Viable System Model (VSM) and Integrated Information Behaviour Model (IIBM) concepts that could be adopted in diagnosing the viability of Functional Analysis of a Gene Sequence (FAoGS) process in a multiple-recursive bioinformatics work situations. The purpose of viability diagnosis is to identify the cohesive and adaptive mechanisms in the organizational structure of FAoGS process. IIBM was the tool used to identify the transformation of one experimental work situation into another. Together with VSM, both concepts help in representing FAoGS process in recursive views.

This paper has explained this basic structure of FAoGS process in terms of the VSM five prescribed functions (*Operational, Coordination, Control, Intelligent, Policy*). Two systemic levels were identified in FAoGS process; meta-systemic and sub-systemic levels which in it contains appropriate VSM functions to diagnose cohesive and adaptive mechanisms. To be able to describe these functions effectively, a model of process recursive levels was produced. This model was represented in terms of transformation of the work situations during FAoGS process. Six recursion levels of the work situations were represented in this model. The basic process structure for Recursion Level 0 (RL0) was presented in this paper and discussed in terms of the five VSM prescribed functions. From the basic process structure, the sub-systemic level shows coordinated information and at meta-systemic level, it shows the situational awareness to certain factors in their information environment. These are the factors that could bring instability and imbalance to the FAoGS process unexpectedly.

ACKNOWLEDGEMENT

I would like to thank Professor Ron Summers at the Department of Information Science, Loughborough University, UK for his supervision in this research study.

REFERENCES

Achard, F., G. Vaysseix, *et al.* (2001). "XML, Bioinformatics and Data Integration." Bioinformatics Review **17**(2): 115-125.

Ahmad, R. and Yusoff, M.B. (2006). A Viable System Approach to Tackle Complex Enterprise Situation for SISP. Malaysian Journal of Computer Science. Vol. 19(1), 87-103.

Ashby, W. R. (1964). An Introduction to Cybernetics. London, Methuen & Co Ltd.

Bartlett, J.C., and Toms, E.G. (2005). "Developing a Protocol for Bioinformatics Analysis: An Integrated Information Behaviour and Task Analysis Approach." Journal of The American Society For Information Science and Technology. Vol. 56(5), 469-482.

Beer, S. (1979). The Heart of Enterprise. Chichester, John Wiley & Sons.

Beer, S. (1985). Diagnosing The System For Organisation. West Sussex, John Wiley & Sons.

Chelsom, J., Katz, S., Zisman, A. and Summers, R. (2002), "Information Bus:Web Services in Action at the United Nations", *XML Conference & Exposition 2002.*

Choo, C. W., B. Detlor, *et al.* (2000). Web Work. Information Seeking and Knowledge Work on the World Wide Web. Dordrecht, Kluwer Academic Publisher.

Detlor, B. (2003). "Internet-based information systems use in organisations: an information studies perspectives." Information Systems Journal **13**(2): 113-132.

Goble, C. (2001). "The low down on e-science and grids for biology." Comparative and Functional Genomics. Vol 2, 365-370.

Espejo, R. (2003). The Viable System Model: A Briefing About Organisational Structure. <http:www.syncho.com/pages/pdf/INTRODUCTION%20TO%20THE%20%SYSTEM %20MODELS3.pdf> [accessed 15:01:06]

Kanehisa. M. (2001). Post-genome Informatics. New York, Oxford University Press.

Systems. A Contextual View. Information Seeking In Context. P. Vakkari, R. Savolainen and B. Dervin. London, Taylor Graham.

Saunders, G.C.(1999), "Quality in the Analytical Molecular Biology Laboratory", in Saunders, G.C. and Parkes, H.C. "*Analytical Molecular Biology : Quality and Validation*", Redwood Books Ltd, Trowbridge, Wiltshire, pp. 14.

Toldo, L. and F. Rippmann (2005). "Integrated Bioinformatics Application for Automated Target Discovery." Journal of The American Society For Information Science and Technology 56(5): 483-492.