

# DATABASE UTILITY IN BIOINFORMATICS -AN EFFORT BASED CURTAIL APPROACH USING MULTI AGENTS

Jayanthi Manicassamy<sup>1</sup>, T. Vengattaraman<sup>2</sup>, and P. Dhavachelvan<sup>3</sup>

Pondicherry University / Department Of Computer Science, India.

Email: {jmanic2@yahoo.com, dhavachelvan@gmail.com}

**Abstract**—Today, in the field of computational biology databases play a vital role due to tremendous incremental in biological datas. The main requirement behinds this is to maintain these biological datas that in turn directly or indirectly utilized by various tools and techniques. These datas plays a major role in accurately solving real time biological problems in the field of bioinformatics. Thus bioinformatics is a means of merging computational biology with few other sciences. In this discipline, there is a need for various discipline precise databases which are most frequently used. Various applications and tools utilize these discipline specific databases for providing various best data solutions. Mostly tools developed in this area involve accessing more than two databases that vary in size and structure. Providing a standard for accessing data's from structural variation databases requires more effort in terms of time, cost and human resources. In this perspective a utility has been developed for reducing the effort for accessing various databases with structural variations. Agents are artificial intelligence capable of sensing the environment and react accordingly to the environment. Based on this aspect, multi agents have been involved in the utility for accessing various databases. In this paper along with the utility developed the novelty explained here is about dynamic accessing of various databases that could be utilized by any tools or applications that requires accessing of various databases. The utility is capable of using any numbers of databases with vary in structure to furnish accurate results which provided a way for reducing efforts that have been visualized based on the analysis made which have also been narrated here

**Index Terms**— Artificial Intelligence, Bioinformatics, Databases, Effort, Evaluation, Multi-Agents, Web-Services.

## I. INTRODUCTION

Biological data's has been predicted in an unpredictable rate for which various databases exist in the field of bioinformatics. Mostly in this area tools and applications developed tools uses more that two databases. A solution has to be provided for problems that are face in extracting the required information's especially in size and structure. For researchers in the biomedical research community for literature search and for analyzing text for contained information where searches are made from bioinformatics databases such as Swiss-Prot, PubMed, MEDLINE, EBIMed etc... with thesauri for data entries and concepts [1-3] [5] for automatic matching and information retrieval [6]. Different tools and techniques have been developed for converting different formatted

data's from various databases to a standard format [10].

In contextual search in biological field, uses specialized search tools that are hosted and utilized to make appropriate search results from various databases. RDF standard helps to make easy access for different formatted databases. In this paper description about the developed multi agent utility has been made in section II along with the mode for embedding the utility with any tools and applications that uses more than two databases of different structure has also been narrated. Reduction of efforts by utilizing utility has been evaluated and the results have been analyzed in section III.

## II MULTI AGENT UTILITY OPERATIONAL METHODOLOGY AND ITS SIGNIFICANCE

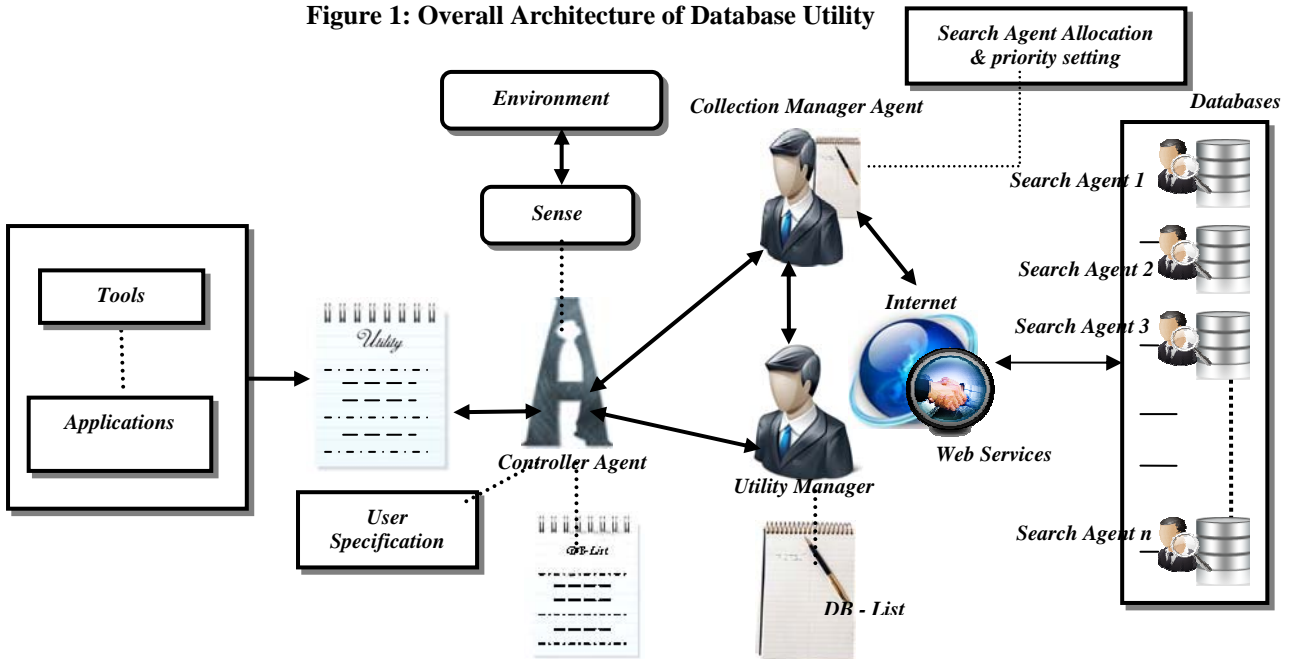
The novel utility for embedding required databases dynamically by using multi agents is capable of embedding with newly developing tools and applications. Based on the primary functionality of the tools or applications databases could be embedded automatically with prioritization or on user defined bases databases will be used with the user prioritization. In automatic embedding of databases with prioritization these activities will be carried out by the agents based on the information extracted from the environment. Various agents are been involved in this activity based on the number of databases number of sub agents are involved to search and receive the appropriate result that could be carried out in parallel which will reduce search time with best appropriate retrieval of results. From retrieved results most appropriate precision will be checked and that particular data's will be considered for further result interpretation.

### A. Operational Methodology

The overall functionality of the utility has been represented in figure 1 of this section which uses multi agents for effective processing. Setting up the databases that is to be utilized by the specific tools or application is based on by invoking the utility, which in turn invokes controller agent which senses the environment and set up the databases that are to be accessed. This agent is the main controller of other agents. The main reason behind the usage of multi agent is to reduce effort and to speed up the process of retrieval ability.

Footnotes: 8-point Times New Roman font; copyright credit, project number, corresponding author, etc.

Figure 1: Overall Architecture of Database Utility



**Controller Agent**

This agent is responsible for controlling and managing all other agents. This agent is invoked by the tools or applications that are been used where, this agent senses the environment is new. If the environmental is new than appropriate databases are searched through collection manager agent and the databases is maintained by this agent. This agent is also responsible for getting search to be made and producing the precised information extracted for further processing by the specific tool or application.

**Collection Manager Agent**

Apart from identifying the specific type of databases for accurate result forecasting, this agent is responsible for dynamically allocating the search agents for each database along with the search to be made. This agent is also responsible for maintaining the precise information extracted by each search agents. The responsibility also lies in selecting the most appropriated information's based on the precision above 95 % is collected the agent decides for the best data selection.

**Utility Manager Agent**

This agent responsibility lies in extracting exact data's that is of highly precised from the collection manager agent based on the list of precise information maintained. Apart from this, it is responsible for specifying the search databases to the controller agent. If the user specific is not set then the list of appropriate databases are retrieved from the controller manager. This task is based on the controller agent where specification are maintained both by the utility manager agent and controller agent. The selection is made for the most appropriate if more than two results found to be of equal precision than the result based on the prioritized search agents will be selected for further processing.

**Search Agents**

These agents are responsible of searching and extracting information along with converting the database data format to a standard RDF format. Here, each search agents are responsible for each database. For the search to be made from the databases parallel search is done by all the search agents for the appropriate results extracting which is maintained by the collection manager agent.

**B. Significance of Multi Agent Utility**

The utility is of efficient and effective usage for reducing work load and time since, there are vast databases each developed for particular data's storage purpose that could not be brought into minds or searched with relevant to the information that is to be processed at an instant. The utility will be capable of automatic connecting to the most appropriate databases with appropriated information extraction accordingly without any human intervention. Parallel searches are carried out by the specific agents in order to reduce search time since, searching the databases in a sequence since vary in database sizes. This utility could be embedded to any tools or any application that is to be developed which performs any biological computational functionality. This utility could be used in the development of effective and efficient tools and applications.

**III EFFORT BASED EVALUATION AND ANALYSIS**

Effort based evaluation has been carried out on tools and applications by embedding this utility in means of cost, resources and time which profound to comparatively more less considering with the tools and applications that doesn't utilizes this utility. Each evaluation and analysis made has been theoretically narrated in this section

**A. Time Based Effort Reduction**

Effort based reduction in means of time have been found more less since the tools and applications embedding this utility doesn't require time in writing a large series of code for convert different structured databases into a

standard format. Apart from this allocating each search agents to each databases and allowing parallel search to all selected databases found to reduce time duration of extracting appropriate information from the large set of databases.

### B. Cost Based Effort Reduction

Effort based reduction in means of cost found to less since the tools and application that utilizes this utility reduces time in means of writing code which in turn reduces the time in utilizing the human resources.

### C. Resource Based Effort Reduction

Effort reduction on resources based in the sense of human resources found to be less compared with the tools and applications that doesn't utilizes this novel utility. Since the utility reduces line of codes for connecting external databases, converting different structured databases to a standards format etc... this in turn reduces human resources

## CONCLUSION

This paper provides details about the novel utility developed for databases that could be embedded to various tools and applications with various functionality which utilize more that two databases to extract information's and to perform their specific functionality. Connectivity to various databases are done based on the functionality appropriate databases are connected on priority bases and information's are extracted for further evaluation to provide an accurate results for real world problems. Multi agents are involves in performing various functionality like searching database, extraction best data's, allocation search agents to each database etc... Parallel searches to various databases are made to extract best appropriate information's. This utility found to be best for utilization in new or existing tools and applications based on the identification and evaluation made based on effort. Apart from this the utility usage is found to be effective and efficient that uses more than two databases for extracting exact information's.

## REFERENCES

- [1] Jayanthi Manicassamy and P. Dhavachelvan, "Metrics Based Performance Control over Text Mining Tools in Bioinformatics", ACM Proceedings, January 2009, pp 171-176.
- [2] Jayanthi Manicassamy and P. Dhavachelvan, "Based Accuracy Perpetuation for Bioinformatics Sequence Analysis Tools", International Journal of Recent Trends in Engineering (IJRTE) - Finland, pp 550-555, May 2009.
- [3] Yoshimasa Tsuruoka, Junlchi Tsujii and Sophia Ananiadou, "FACTA: a text search engine for finding associated biomedical concepts", ACM Portal, September 2008.
- [4] Dean Cheng, Craig Knox, Nelson Young, Paul Stothard, "PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites", Oxford Journals, pp 399-405, May 2008.
- [5] Weijian Xuan, Pinglang Wang, Stanley J. Watson and Fan Meng, "Medline search engine for finding genetic markers with biological significance", ACM Portal, pp 2477-2484, July 2007.
- [6] Qing Xiong, Yuhui Qiu and Weikuan Gu, "PGMapper: a web-based tool linking phenotype to genes", ACM Portal, pp 1011-1013, Jan 2008.
- [7] Matteo Fattore and Patrizio Arrigo, "Knowledge Discovery and System Biology in molecular medicine: an application on neurodegenerative diseases", ISB, December 2004.
- [8] Raman Karthik and Chandra, Nagasuma, "PathwayAnalyser: A Systems Biology Tool for Flux Analysis of Metabolic Pathways", Nature preceding, May 2008.
- [9] Michael Baitaluk, Mayya Sedova, Animesh Ray and Amarnath Gupta, "BiologicalNetworks: visualization and analysis tool for systems biology", Oxford Journal, pp W466-W471, March 2006.
- [10] Francois Belleau, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault "Bio2RDF: Towards a mashup to build bioinformatics knowledge systems", Elsevier, September 2007.
- [11] Rodrigo Santamaria, Roberto and Luis Quintales, "BicOverlapper: A tool for bicluster visualization", ACM Portal, pp 1212-1213, March 2008.
- [12] Xiang Xu, Jinyu Wu, Jian Xiao, Yi Tan, Qiyu Bao, Fangqing Zhao and Xiaokun Li, "PlasmoGF: an integrated system for comparative genomics and phylogenetic analysis of Plasmodium gene families", ACM Portal, pp 1217-1220, March 2008.
- [13] Juby Jacob, Marcel Jentsch, Dennis Kostka, Stefan Bentink and Rainer Spang, "Detecting hierarchical structure in molecular characteristics of disease using transitive approximations of directed graphs", ACM Portal, pp 995-1001, February 2008.
- [14] Maryan E.Martone, Joshuna Tran, Willy W. Wong, Stephen Larson, Stephan P. Lamont, Amarnath Guptha, Mark H. Ellisman, "The Cell Centered Database Project : An updated on building community resources for managing and sharing 3D imaging data", Elsevier, October 2007.
- [15] Ana Carmena, "Signaling networks during development: the case of asymmetric cell division in the Drosophila nervous System", Elsevier, June 2008.
- [16] Francois Belleau, Marc-Alexandre Nolin, Rigault, Jean Morissette, "Bio2RDF: Towards a mashup to build bioinformatics knowledge systems", Elsevier, September 2007.
- [17] Yang Jin, Ryan T McDonald, Kevin Lerman, Mark A Mandel, Steven Carroll, Mark Y Liberman, Fernando C Pereira, Raymond S Winters, Peter S White, "Automated recognition of malignancy mentions in biomedical literature", November 2006
- [18] Bilgin, C. Demir, C. Nagi, C. Yener, B., "Cell-Graph Mining for Breast Tissue Modelling and Classification", IEEE , pp 5311-5314, August 2007.
- [19] James V. Freeman, Frederick E. Dewey, David M. Hadley, Jonathan Myers, and Victor F. Froelicher, "Autonomic Nervous System Interaction with the Cardiovascular System during Exercise", Progress in Cardiovascular Diseases, Vol. 48, No. 5 (March/April), 2006: pp 342-362.
- [20] Rayavarapu Bhargavi, U. Suryanarayana Murthy and G. Narahari Sastry, "Structural and active site analysis of plasmepsins of plasmodium falciparum: Potential anti-malarial targets", Elsevier, August 2005.