User-Centered Design and Evaluation of a Dynamic Biochemical Pathway Visualization Tool

By

Rana Khartabil

B. Sc. Concentration in Biochemistry (1998), B. A. Sc. in Software Engineering (2003)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master in Computer Science

Ottawa-Carleton Institute for Computer Science

School of Information Technology and Engineering

The University of Ottawa,

Ottawa, Ontario, Canada

Acknowledgements

This research is supported by the Institute für Zell- und Organ Simulation (IfCOS), a biotechnology company based in Rostock, Germany.

I would like to express my deepest gratitude and thanks to the following people who have contributed to my research:

- My supervisor, Dr. Timothy Lethbridge, for his knowledge, good advice and guidance throughout my research.
- My co-supervisor, Dr. Leonard Kleine, for his feedback on biochemistry related issues and for supplying me with the contacts and users I needed to help me with my research.
- All members of the thesis defense committee for their valuable feedback.
- Dr. Stephan Gutjahr, the CEO of IfCOS, who has inspired me to conduct this research, and who believed in my capabilities.
- All colleagues of IfCOS, including Dr. Stephan Gutjahr, Dr. Karin Schulz, Manfred Pruntsch, Hartmut Reichstein, Loay Khartabil, and Arif Rajwani, for their contribution.
- Members of the Knowledge Based Reverse Engineering (KBRE) group, who have given me some valuable comments and feedback on my research.
- All users who have participated in the user studies for their time and valuable feedback.
- My family and friends for their love and support.

Abstract

Information visualization is a field of computer science that deals with the computerized visualization of complex information in a form that is easier for human beings to comprehend. Information visualization has applications in many domains, including business, science, and medicine. Visualization of biochemical pathways, as graphs of nodes representing biochemical entities and arcs representing the relationships between entities, is one such application.

This thesis begins by reviewing work that has been done on the usability of information visualization techniques, and in particular these that apply to biochemical pathways. Then, the thesis presents three different usability evaluation techniques that are used to gather information about existing biochemical pathway visualization tools. These are (1) conducting videotaped evaluation sessions of existing biochemical visualization tools, (2) collecting questionnaires, and (3) conducting a brainstorming session. The results from these studies are used to define the requirements, design, and build a biochemical pathway visualization tool, taking into account conclusions drawn from both literature and user studies. The tool is then tested and compared to existing tools.

Results show that the developed tool has more relevant features to biochemical pathway visualization than existing tools, accomplishes certain tasks faster than other tools, and is intuitive and easy to use. In addition, positive feedback from users is documented.

At the end of the thesis, we make some generalizations to the area of information visualization and we then present areas for further research.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction to Biochemical Pathway Visualization

The purpose of this research is to design an easy to use user interface (UI) for biochemical pathway visualization and to evaluate the usability of this UI. The methodology used to get input on how to design such a UI consists of reviewing previous work on the usability of information visualization, in particular biochemical pathways, and conducting usability experiments to collect data on user preferences. The usability experiments consist of conducting a videotaped evaluation of existing tools, collecting questionnaires on user preferences, and conducting a brainstorming session. The input from these methods is used to design the UI of BioPathVis, our developed biochemical pathway visualization tool. Later, BioPathVis is evaluated using a videotaped evaluation and a follow up questionnaire. The methodology and results will be described in details later in the thesis.

This research has shown a number of fundamental contributions, which will be demonstrated throughout this thesis. Many of these contributions would be useful to people doing any kind of information exploration, and most ideas would be useful for exploration of complex data that is in graphical form. The first contribution is the methodology itself, which has shown to be valid and effective in designing an intuitive, comprehensive, and easy to use tool.

The second contribution is the biochemical pathway visualization tool, BioPathVis. While designing the UI of BioPathVis, we were faced with several choices for designing certain UI features. The resulting tool encompasses the following features, which synergistically combine with each other:

- The use of a tree view to display pathway graphs.
- The use of a tabbed panel to display static overview images as well as dynamic individual pathway graphs.
- The use of static KEGG images to display overview images of pathways.

- The ability to save images of pathways displayed, so they can be imbedded in other documents.
- Various search capabilities. We decided, after analyzing the alternatives, to include pathway organization search (species, organ, tissue, cell, and organelle) in a quick search feature and additional capabilities in a detailed search feature.
- Displaying a legend.

A detailed explanation of the design decisions leading to the above will be presented later in the thesis.

The third contribution is pointing out flaws in other systems, through a malfunction analysis using the videotaped evaluation results. The malfunction analysis helped us to determine which practices to be followed or avoided. In addition, implementers of existing tools can use these results to further enhance their tools, by considering the recommendations for change shown in the appendix.

It is worth mentioning that some of the malfunctions and design decisions presented in this thesis are related largely to 'utility' (i.e. functionality), whereas others are related largely to 'usability'. The latter results from not adhering to the UI guidelines, whereas the former result from deeper analysis.

1.1 What are Biochemical Pathways?

Biochemical pathways are interconnected networks of molecular reactions, interacting under given physiological conditions via simple intermediates [16]. Biochemical pathways represent important aspects of cellular processes of living organisms, including metabolism and regulation.

1.2 Biochemical Pathway Classification

Biochemical pathways are categorized into metabolic pathways and regulatory pathways. Metabolic pathways describe all chemical reactions that occur in living cells, including those that degrade complex molecules to liberate smaller molecular building blocks and energy (catabolic reactions) and those that synthesize molecules needed for cell maintenance and growth (anabolic reactions) [28]. Enzymes facilitate most of these chemical reactions and thus pathways containing such reactions must not only describe the substrates and the products of the reactions, but also the enzymes that catalyze these reactions. Metabolic pathways are most commonly classified according to the metabolism of the molecules, such as carbohydrates, lipids, nucleotides, amino acids, cofactors, vitamins, etc. Please refer to Table A-1 of Appendix A for a list of known metabolic pathways, classified by molecules metabolized. This table is taken from KEGG (Kyoto Encyclopedia of Genes and Genomes) [20].

Regulatory pathways describe biological processes that are controlled by different signals. These include genetic information processes, such as transcription and translation, cellular environmental information processes, such as membrane transport, signal transduction, and ligand-receptor interaction, cellular processes, such as cell mobility, cell growth and death, cell communication, development, and behavior. Please refer to Table A-2 of Appendix A for a list of known regulatory pathways, classified by biological processes. This table is also taken from KEGG [20].

1.3 Application of Information Visualization to Biochemical Pathways

Information visualization is a field of computer science that deals with the computerized visualization of complex information in a form that is easier for human beings to comprehend. Information involved in biochemical pathways is one example of complex information that is well suited for information visualization. Looking at the metabolic

network of the organism E. coli helps to illustrate the complexity of biochemical networks. E. coli alone contains 791 chemical compounds organized into 744 enzyme-catalyzed biochemical reactions [18]. Such complexities are best dealt with and understood by using computer databases and software to visualize this information.

In the field of information visualization, when relationships exist between the data elements to be visualized, one can represent the information as a graph of nodes representing entities and arcs representing the relationships between entities [34]. Since biochemical pathways represent biochemical processes in the form of complex, interconnected networks of biochemical reactions, biochemical pathway visualization is one application of information visualization using graphs.

Visualization of biochemical pathways allows for the storage of current knowledge of all biochemical pathways for organisms and provides biochemists with a more intuitive understanding of the relationships between the various compounds involved in biochemical pathways. This aids biochemists in making new biological discoveries; as such knowledge can be used in predicting new pathways, discovering or inventing drugs, and in the search of possible causes of genetic diseases. Information on the metabolic pathways of one organism can help understand the metabolic pathways of a newly sequenced genome. In addition, studying the metabolic and regulatory pathways allows scientists to understand abnormalities that are the causes of diseases, which is an important step in drug discovery. Also, computer simulation of regulatory networks allows for the identification of “candidate genes” that are responsible for diseases [36]. Once the cause of the disease is known, one attempts to develop a drug that will minimize or remove the cause. Simulation of such networks with the new drug and studying the influence of the presence of the drug in the organism helps in understanding the possible side effects, which is another important step in drug discovery.

Thus, computerized tools that aid in biochemical pathway storage, retrieval, manipulation, and visualization, are needed. However, the complexity and the vast amount of data involved in these pathways makes building such tools a very complex task. Nevertheless, several computerized systems for biochemical pathway visualization and analysis were built, such as

BioCarta, ExPASy, KEGG, MetaCyc, WIT, etc. Several of these systems are available for free for academic use; others are only available for commercial use. Below is a brief overview of some common biochemical visualization tools.

1.4 Overview of the Most Common Biochemical Pathway Visualization Tools

1.4.1 BioCarta

BioCarta is a web-based tool that allows for the visualization of both metabolic and regulatory pathways as static drawings. The main focus of BioCarta is to create tools to study pathways [3]. BioCarta allows the user to search for pathways by their name or class (i.e. Metabolism, Cell Signaling, etc), and to search for compounds. Figure 1-1 below shows the Glycolysis pathway representation in Homo sapiens, using BioCarta. In the pathway figure, substrates of chemical reactions are shown in their chemical form, with their names at the bottom of the chemical structure, and enzymes are represented as numbers, with a legend at the bottom specifying their names. Only the enzymes in the figure are clickable, and clicking on enzymes displays information about them. The visualization portion of BioCarta is free, whereas the editing portion requires a license.

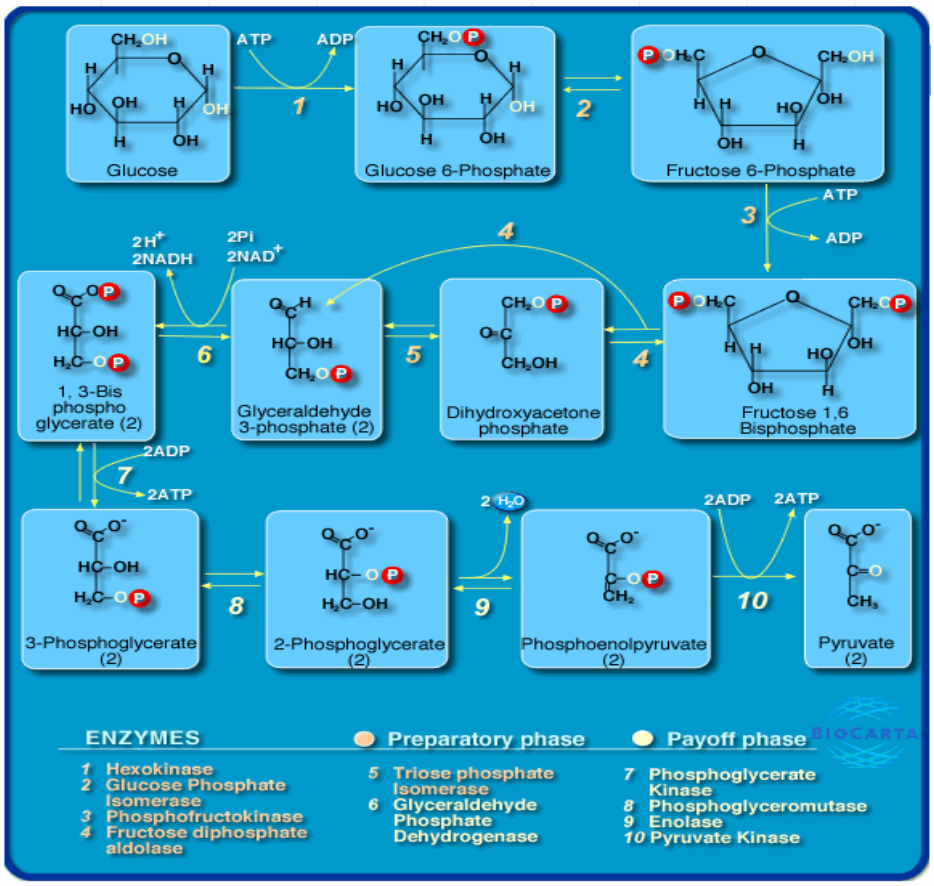


Figure 1-1: Glycolysis pathway representation using BioCarta [3].

1.4.2 ExPASy

The ExPASy (Expert Protein Analysis System) Molecular Biology Server [10] is a tool that is focused on the analysis of protein sequences and structures [10]. The pathway tool of ExPASy is a web-based static biochemical pathway visualization system, which gives direct access to the scanned-in version of the Boehringer Mannheim “Biochemical Pathways” map [27]. This map is partitioned into 115 pieces and results of the queries are returned by matching keywords against the entries in the map [3]. ExPASy allows the user to search for a pathway map, given a keyword that is present in the map. It does not allow the user to search for pathways by name or class, nor does it allow the user to add or edit pathways. It also does not allow the user to search for or display information about compounds.

Figure 1-2 below shows the Glycolysis pathway representation using ExPASy. Only enzymes in the figure are clickable to display further information. Arrows outside the picture display the linked pathways. For a complete legend, please refer to Michal's "Biochemical Pathways" [28]. For Figure 1-2 , it is sufficient to know the following [28]:

- Substrates of enzymatic reactions are shown in black, enzymes in blue, and coenzymes in red.
- Orange is used for regulatory effects.
- The color of the reaction arrow shows where the reaction was observed: black represents general pathways, blue represents pathways observed in animals, green represents pathways observed in plants and yeast, and red represents pathways observed in prokaryotes.
- Bold arrows indicate main pathways of metabolism.
- Points on both ends of an arrow indicate reversibility of this reaction under biological conditions.

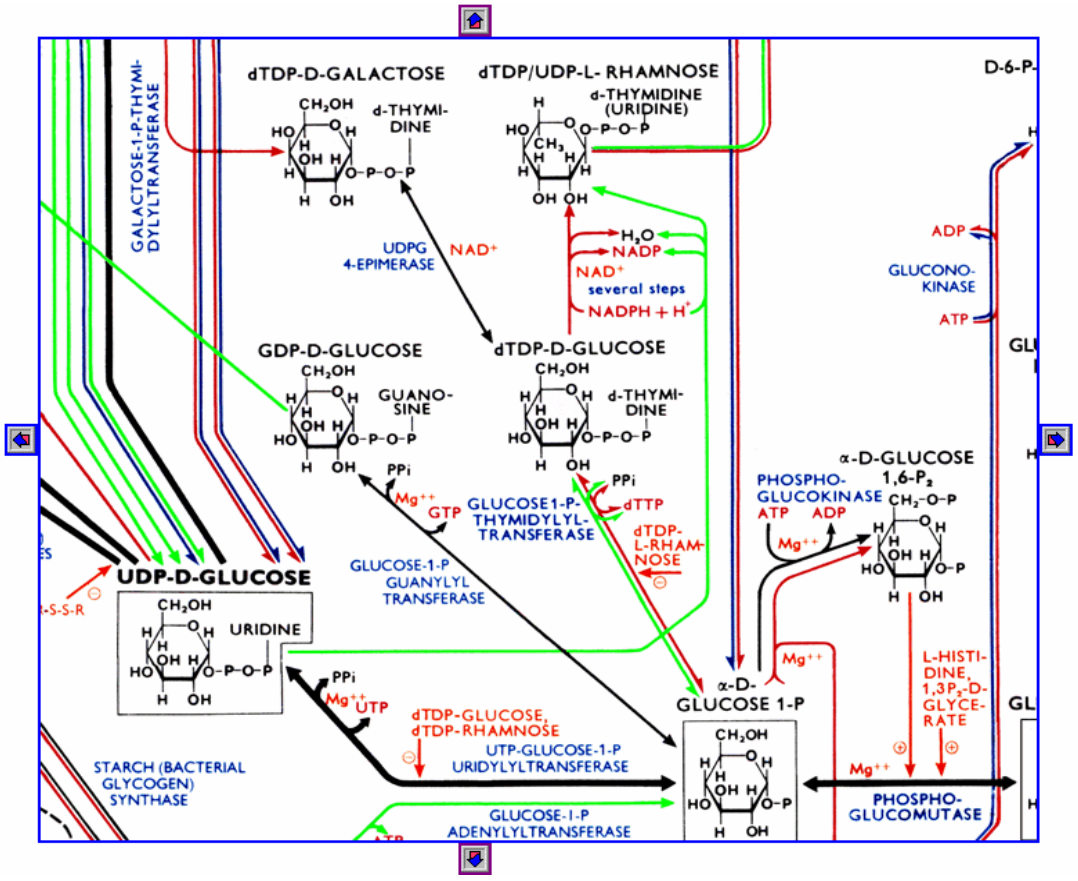


Figure 1-2: Glycolysis pathway image map using ExPASy [10].

1.4.3 KEGG

The Kyoto Encyclopedia of Genes and Genomes (KEGG) system is a web-based suite of databases and software that aims to combine the current knowledge of genetics, biochemistry, molecular and cellular biology, and to visualize this knowledge as pathways of interacting molecules or genes [17]. KEGG is composed of 3 databases: The PATHWAY database, which consists of graphical diagrams of most metabolic pathways and some of the regulatory pathways, the GENES database, which contains gene catalogues of all organisms with completely sequenced genomes and some organisms with partial genomes, and the LIGAND database, which contains information about the various biochemical compounds [17].

KEGG visualizes pathways statically. It allows the user to search for pathways in different organisms by their class and name, as well as to search for compounds. The PATHWAY database contains GIF image maps of biochemical pathways that are manually drawn and continuously updated [17]. These maps allow for the display of additional information on compounds and enzymes and the display of adjacent pathways, by clicking on these compounds or pathways. Figure 1-3 below shows an example of the KEGG image map of the Glycolysis pathway for E. coli. Substrates and products of biochemical reactions are drawn in circles, enzymes, represented by their Enzyme Commission (E.C.) number, are drawn in rectangles, and adjacent pathways are drawn in semi-rectangles. Enzymes found in the gene catalog of a specific organism are marked in green [17]. Clickable items in this diagram are the enzymes, metabolites, and adjacent pathways.

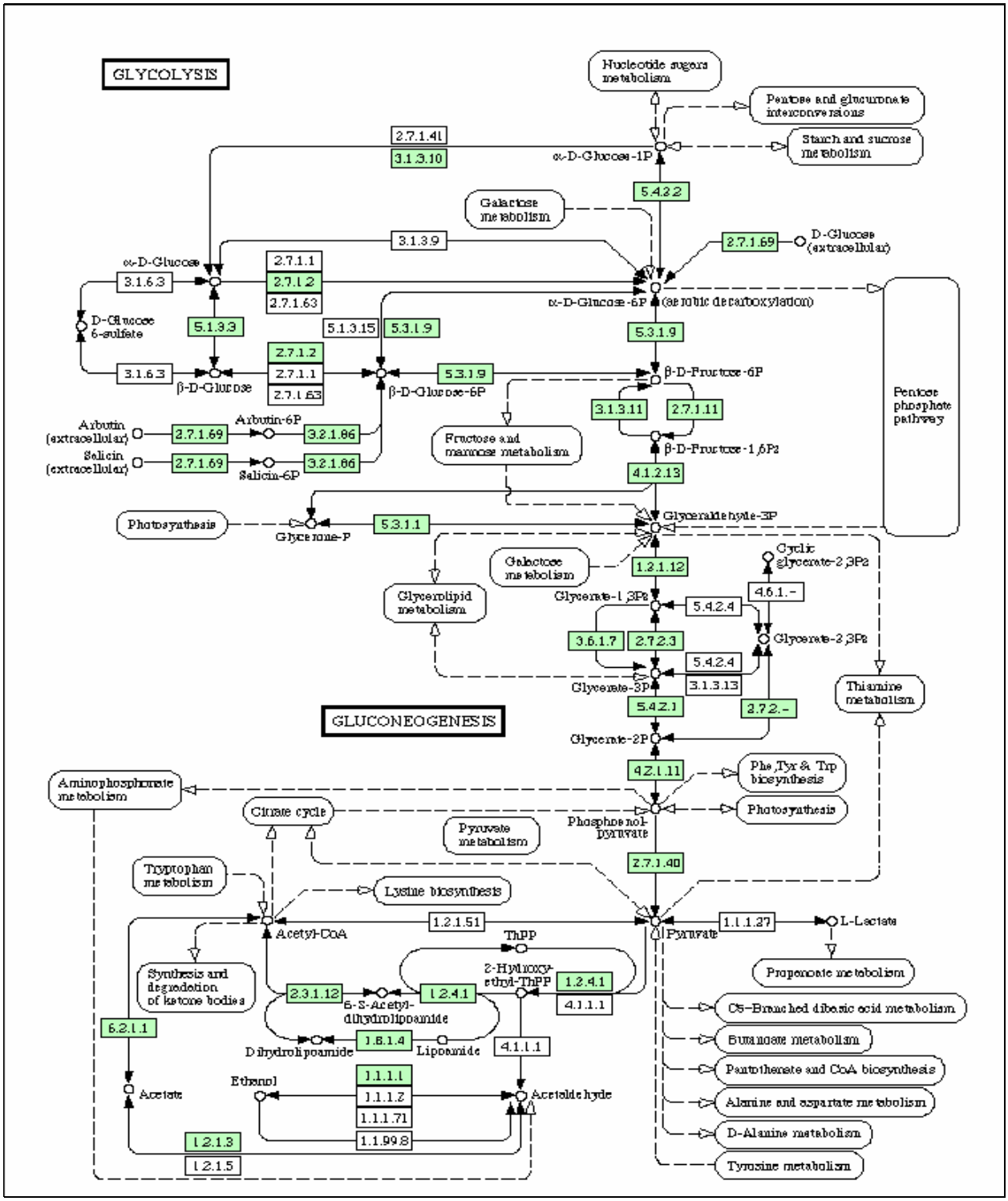


Figure 1-3: Glycolysis pathway representation in E. coli using KEGG [17]

1.4.4 MetaCyc

MetaCyc is an electronic encyclopedia, of over 450 metabolic pathways from over 150 different organisms, that allow scientists to explore genomic and biochemical information ([18], [19], [26]). It consists of a knowledge base, which describes the genes and compounds, and a graphical user interface to access that knowledge [19]. The software is implemented in COMMON LISP. The reason that Artificial Intelligence techniques are incorporated in MetaCyc is to allow inference, such as pathway prediction from sequenced genomes ([18], [19]). MetaCyc uses a graph layout algorithm for drawing metabolic pathways dynamically at run-time, which reads metabolic pathway information from a database, breaks up the metabolic pathway graph into cyclic, linear, and tree-structured components, and then applies different layout methods to each [3]. The user interface of MetaCyc allows the user to search for pathways in different organisms by their class and name, as well as to search for compounds. It also provides hypertext navigation among related entities and allows for showing more or less detail of the viewed pathways [19]. Figure 1-4 below shows an example of the MetaCyc of the Glycolysis pathway for E. coli. All items in the figure, including the reaction arrows, are clickable to display further information.

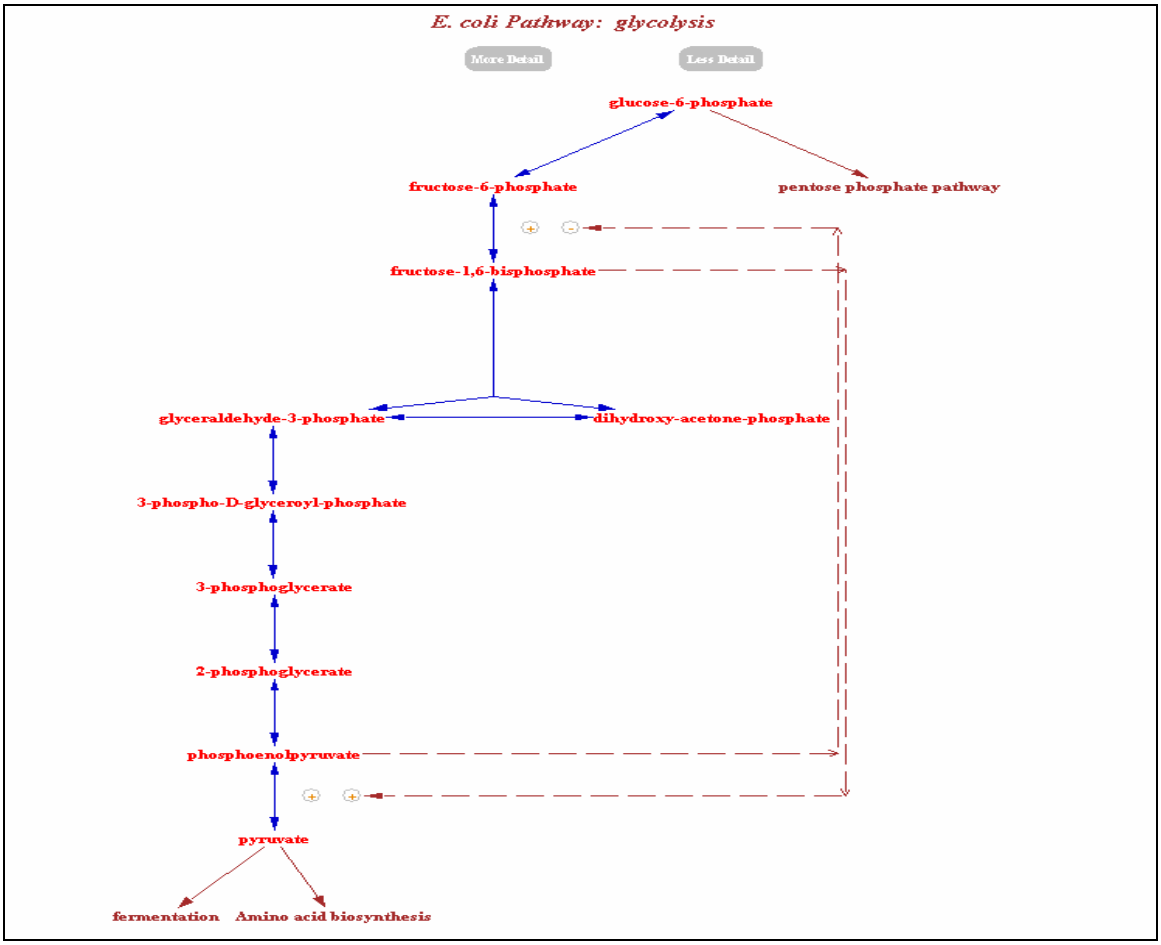


Figure 1-4: Glycolysis pathway representation in E. coli using MetaCyc [26]

1.4.5 WIT

The WIT (What is There) system [40] was designed to allow comparative analysis of sequenced genomes and to generate metabolic networks based on genetic and metabolic data from the EMP (Enzyme and Metabolic Pathways) and MPW (Metabolic Pathways) databases ([16], [33]). The pathway tool of WIT is web-based and it presents static views of metabolic pathways. It allows the user to search for pathways in different organisms by the metabolites and enzymes that are present in the pathway, and not by the pathway class or name. Figure 1-5 below shows an example of the WIT representation of the Glycolysis pathway for E. coli. Enzymes and metabolites in the figure are clickable, and clicking on them displays further information.

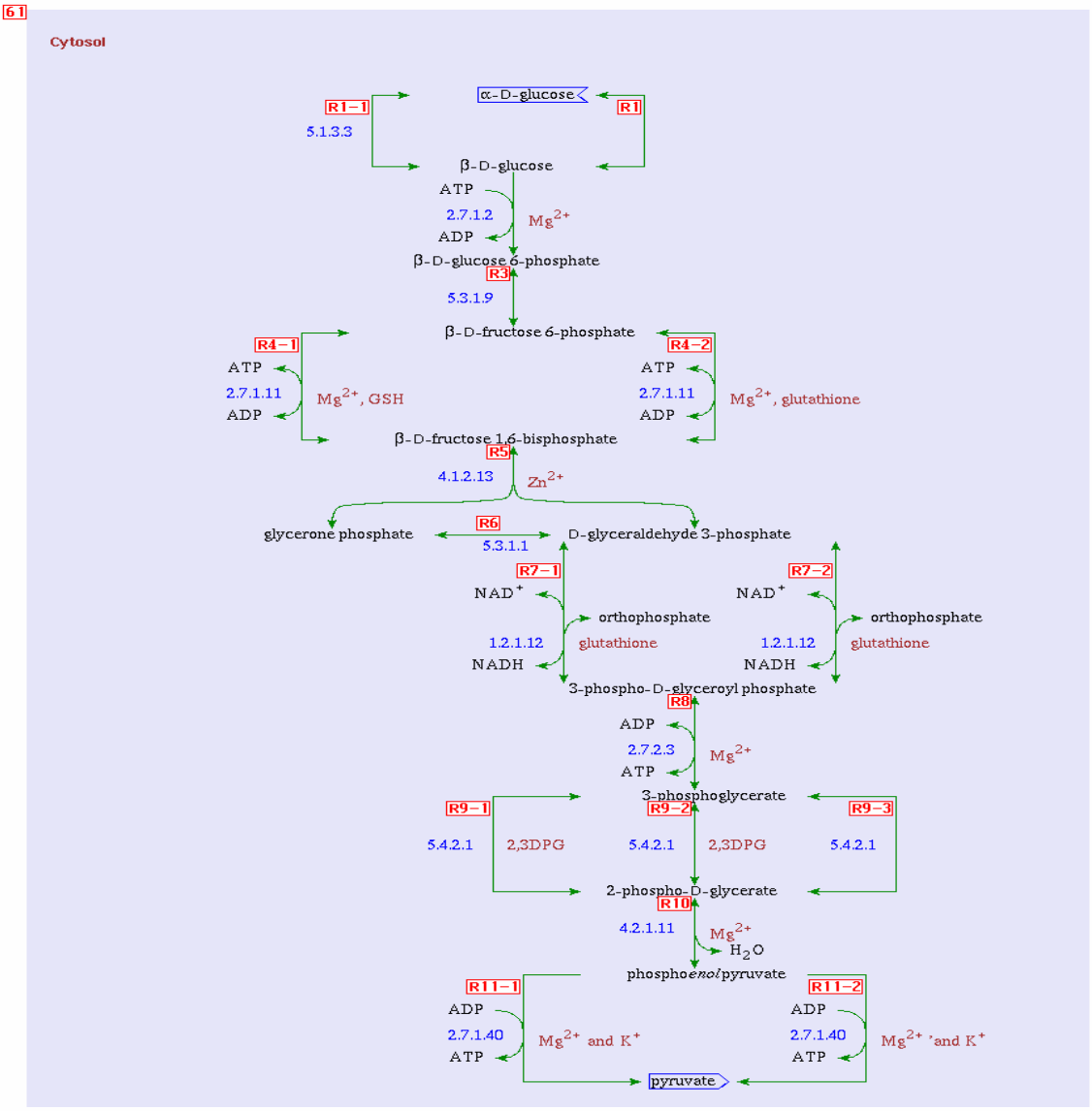


Figure 1-5: The WIT image map of the Glycolysis pathway in E. coli [40].

1.4.6 Other Tools

There exist some other tools, which are commercially available and do not solely focus on biochemical pathway visualization. Some of these tools are briefly discussed below.

1.4.6.1 Visual Cell

VisualCell is a commercial tool, built by the Gene Network Sciences (GNS) company. The purpose of GNS is to provide tools to accelerate the drug discovery process by creating dynamic computer models of living cells [12]. VisualCell is used to build pathways, store information, overlay expression data and manage knowledge using a language, called the Diagrammatic Cell Language (DCL) [12]. DCL enables users to represent thousands of bio-molecules and their interactions (such as signal transduction pathways, gene expression networks and metabolic pathways) and then quantitatively simulate the model [12]. The pathway visualization tool of VisualCell has its own modeling language symbols and icons that are different from the standard pathway visualization tools.

1.4.6.2 PhysioLab

PhysioLab is another commercial tool that is built by Entelos. Entelos provides tools and services to develop dynamic, large-scale computer and mathematical models of human diseases, which helps pharmaceutical and biotech organizations to develop effective new treatments for diseases more rapidly [9]. PhysioLab can be used to: (1) identify and characterize novel pathways and genes, (2) prioritize and evaluate targets and candidates, (3) plan and optimize clinical trials and experiments, (4) assess the clinical impact of therapeutic approaches on different patient populations, and (5) relate genomic, proteomic, and *in vitro* data to clinical outcomes [9].

1.4.6.3 PathwayPrism

PathwayPrism™ is a web-enabled, customizable commercial tool that integrates pathway analysis and simulation [35]. It is built by Physiome Sciences, which is another company that aims at helping pharmaceutical companies to develop drugs faster through the use of biological simulations in the form of mathematical models. PathwayPrism™ is used to map, analyze and simulate molecular interactions in the cell [35]. The biochemical pathway visualization portion of the tool can be used to create and merge complex pathway maps.

1.5 Motivation for Developing New Biochemical Pathway Visualization Tools

Most of the biochemical pathway visualization tools that exist today, such as BioCarta, ExPASy, and KEGG, use static pages for viewing pathways. MetaCyc is one of the first developed tools to view pathways dynamically at run time [3]. While static representation of pathways is easier to implement and presents similarities to biochemistry textbooks, it imposes certain difficulties. The first is that images have to be edited manually whenever data has to be updated; the second is that there is no way to specify details or hide parts of pathways; and the third is the inapplicability of static visualization when it comes to visualizing novel or user-defined pathways [3]. On the other hand, dynamic representation of pathways, despite its complexity, provides the high flexibility necessary for complex queries and for the construction of new pathways [3]. For these reasons, the biochemical pathways visualization process is best accomplished dynamically at run-time, based on the information read from the database.

Although MetaCyc and some other dynamic visualization tools have solved the problem of static visualization, they still have their own shortcomings. The first is the lack of pathway editing capabilities in the software itself, under Windows. For users to add or edit new pathways, users must manipulate the database directly or use additional Editing tools under platforms other than Windows. It would be more efficient and convenient if users can edit or add new pathways directly from the user interface, under all platforms, in a manner that updates the database automatically.

Usability is a term used to describe the quality and ease of use of an application by users [24]. It is aimed at identifying problems, which interfere with the ability of the user to accomplish certain tasks, and identifying possible solutions to these problems [24]. Paying proper attention to usability will, among other things, allow for a balance between the complexity of a tool and the ease of use of its features. We have not seen any records of usability and human factor issues considered while building biochemical pathway

visualization tools. As such, most of these tools are difficult to use and might require reading a manual or consulting help pages before usage. User preferences must be considered when building these tools. For instance, questions such as which biochemical pathway representation is the best for users, which navigation and interaction methods are most effective and easier to use, which functionalities are desired, etc, must be answered, by conducting user interviews and experiments.

Another interesting area related to usability is graph layout. Although many papers have discussed 3D representation in information visualization ([5], [6], [13], [34], [37]), none have considered building a tool to visualize biochemical pathways in 3D and study the usability of such representation.

Thus, new methods are needed to visualize and manipulate biochemical pathways in a dynamic, more efficient, and more effective manner, while taking into account user preferences and human factors' considerations. The usability issues that must be considered when dealing with information visualization using graphs are mostly related to graph layout and 3D navigation and interaction mechanisms.

Chapter 2 Usability and Human Factors' Considerations in Biochemical Pathway Visualization

2.1 General Criteria for Assessing the Usability of Visual Representation and Interaction Mechanisms

Some work has been conducted to define general criteria for assessing the usability of information visualization techniques. Luzzardi et al established criteria for evaluating the usability of visual representation and interaction mechanisms as a first step in evaluating information visualization [24]. These criteria are described in Table 2-1.

Table 2-1: Criteria for evaluating the usability of visual representations and interaction mechanisms (Paraphrased from [24])

Technique	Criteria	Description
Visual Representation	Limitations	Defines geometric or visual constraints (e.g. size of display).
	Cognitive Complexity	Measured by data density (e.g. number of points in graph), data dimension (e.g. number of dimensions displayed simultaneously), and the relevance of displayed information.
	Spatial Organization	Measured by the overall layout, how easy it is to locate information elements, and their spatial orientation.
	Information Coding	Defines the mapping of data to visual elements and the use of symbols to aid in perceiving information.

Technique	Criteria	Description
	State Transition	Defines rebuilding of visual representation after user actions, including the time to rebuild the representation and changes in spatial organization.
Interaction Mechanisms	Orientation and Help	Measured by allowing users to control levels of detail, undo capability, and the presence of help.
	Navigation and Querying	Measured by the easiness of selecting data elements, changing user's point of view, changing geometric representations of data elements, expanding or hiding data elements, and searching for specific information.
	Data Set Reduction	Measured by providing filtering (reduction of information), clustering (representing subsets of data by special symbols), and pruning (cutting off irrelevant information from display).

The above general criteria must be taken into consideration when designing and evaluating the usability of biochemical pathway visualization tools. In particular, the visual representation of biochemical pathway visualization tools must take into account the following:

- Limitation: The size of display for biochemical pathways is defined by the size of the computer screen, and also by the size of the 'virtual canvas' area to which the user can scroll.
- Cognitive Complexity: The data density of biochemical pathways can be up to tens of thousands of compounds and reactions, if all the pathways are displayed simultaneously, and up to hundreds of compounds and reactions, if the pathways are displayed separately.
- Spatial organization: Information elements should be easily located on the pathway figure. The overall layout must be easy to read.
- Information coding: Data contained in biochemical pathways must be mapped to visual elements. For instance, pathway data read from the database must be mapped to pathway objects containing attributes. Examples of such objects are pathway

objects, reaction objects, and compound objects. Also, such objects must be easily distinguishable. For instance, different shapes and color can be used to represent different objects.

- State transition: Examples of applications of state transition for biochemical pathways occur when the user zooms in and out, or when the user hides or shows additional information on the pathway figure. In such tasks, the time to rebuild the pathway diagram must not be large and the spatial organization of pathways must not change.

In addition, the interaction mechanisms in biochemical pathway visualization tools must take into account the following:

- Orientation and Help: The user must be able to control the level of detail by showing or hiding information on the pathway figure. Help and tool tips on using features in the tool must also be provided, especially for complicated tasks.
- Navigation and querying: Data elements in the pathway figure must be easily selected. It should also be easy to hide and show information. In addition, general and detailed searching must be provided.
- Data Set Reduction: Some sort of filtering and/or clustering of information must be provided.

The subsequent sections describe more specific considerations for biochemical pathway visualization.

2.2 Graph Layout

2.2.1 Introduction

Since biochemical pathways are typically modeled as graphs of nodes, representing chemical compounds, and arcs, representing chemical reactions, biochemical pathways visualization

can be thought of as a graph layout problem [3]. Graph Layout algorithms are concerned with calculating the positions of the nodes of the graph and the arcs connecting these nodes.

When visualizing information as graphs, the most important criteria that influences design decisions in determining the layout algorithm is the size of the graph. A large graph reduces performance and imposes restrictions on the amount of information that can be viewed [13]. Herman et al. stated that users comprehend smaller graphs better, and that a certain layout algorithm may produce good layouts when the size of the graph is less than several hundred nodes [13]. However, graphs in information visualization may contain more than tens of thousands of nodes, such as in the case of biochemical pathways. Many efforts have been made to visualize huge graphs in a way that is easy to comprehend. The next section presents one proposed solution: using 3D.

Other important usability criteria that must be considered in graph visualization are predictability and time complexity. Predictability means that the layout algorithm with similar graphs must lead to similar visual representations, while time complexity means that the visualization system must provide near real-time interaction [13].

2.2.2 2D vs. 3D

Many information visualization systems employ 2D layout algorithms. When dealing with 2D graphs, certain aesthetic rules must be applied to enhance the usability of the visualization to the end-user. Out of these rules, the most relevant to the end-user and the one that has precedence over all other rules (such as maximizing symmetry) is minimizing edge-crossings [13].

All of the tools that deal with biochemical pathway visualization only employ 2D layout algorithms. For example, the graph layout algorithm used in MetaCyc breaks the graph into cyclic, linear, and tree-structured components [3]. Another algorithm for drawing biochemical pathways was proposed by Becker et al. is force-directed to ensure planarity,

minimal edge crossing and drawing area, and maximal symmetry [3]. Other methods involve modeling nodes and edges of the graph as physical bodies tied with springs and aim at minimizing edge crossings [13].

As was mentioned, 2D graphs impose restrictions when the size of the graph is large. These restrictions are associated with the amount of information that can be visualized. To solve the problems associated with visualizing large graphs in a manner that is easily perceived and understood by users, some papers have suggested the use of different pages or windows ([2], [34], [38]), as will be discussed in section 2.3.2.4. However, most papers on information visualization provide 3D graph layout as the ultimate solution to visualizing large graphs. Most argue that 3D is better than 2D ([5], [6], [13], [34], [37]), since the extra dimension gives more space, allowing the display of larger amounts of information and the creation of real-world metaphors, and allowing the user to navigate better. Another paper states that using stereo and motion depth cues in 3D can provide 3 times more information [38]. In addition, minimizing edge-crossings is not required in 3D, since edge crossings can be resolved in depth in 3D and because edges are less likely to intersect in 3D ([34], [37]).

However, 3D visualization has its own difficulties. These are mostly related to the cognitive aspects of 3D navigation, namely the perceptual and navigational conflicts that are caused by the discrepancy of using 2D screens and input devices to interact with a 3D world, combined with missing motion and stereo cues ([13], [31]). In addition, despite the fact that 3D entities look more realistic, unlike 2D entities, they should be equally understandable from any viewpoint [34]. Also, one paper states that the "amount of additional semantic information that can be conveyed by a three-dimensional solution is outweighed by an associated increase in cognitive demands", and thus favoring 2D layouts to 3D [5]. Further, Neilson lists some additional difficulties related to time consumption of 3D graphs and to the poor screen resolution that makes it impossible to render objects in the background in sufficient detail to be recognizable by users [31]. Neilson suggests that 3D should not be used to represent abstract information or to navigate through hyperspace and that 3D graphs should only be used to visualize actual physical objects that can be better understood in 3D, such as molecules, the human body, etc., abstract data types with three attributes, and video games

[31]. Furthermore, experiments have been conducted to compare 2D image browsers to 3D. The results suggest that 2D browsers are more effective than 3D for larger image set sizes [38].

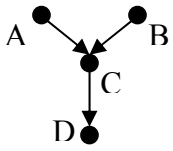
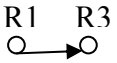
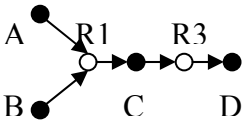
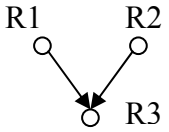
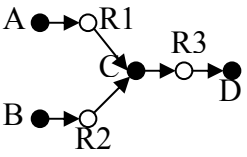
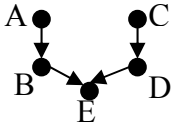
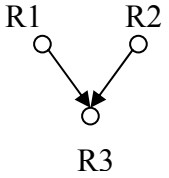
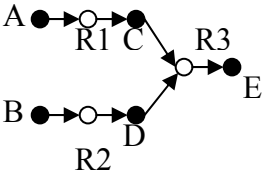
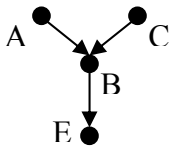
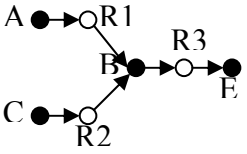
The layout algorithms in 3D consist of several stages: making the graph acyclic, assigning nodes to layers, and assigning nodes within the individual layers [37]. An example of the 3D graph layout techniques is cone trees, which is developed directly for 3D and which allows the user to pick any node and rotate it in the tree to bring it to front [13]. An example of 3D data visualization software is NestedVision3D (NV3D). It uses 3D widgets, rapid zooming, and interactive elision (See section 2.3.2) for navigating graphs with greater than 35,000 nodes and 100,000 relationships [34]. Another example is SemNet, which includes a variety of strategies for 3D automatic layout of graphs, including spring forces and simulated annealing [37].

Despite the fact some papers on information visualization have mentioned the possible application of 3D to drawing biochemical pathways, such as the paper by Auber [1], none have actually constructed a tool that focuses solely on visualizing biochemical pathways in 3D. More usability analysis and comparison based on experiments must be conducted in order to make the decision as to which is better in terms of usability. At the current time, since no experiments have been done to show which technique is better for biochemical pathways, the best approach, which is the one that will be used to construct the biochemical visualization tool, is to allow for a mix of 2D visualization and 3D navigation and interaction techniques. In fact, one paper discusses that "3D interactive techniques might best be introduced alongside more familiar 2D visualizations, allowing the user to mix interaction strategies as necessary" [5]. In addition, one paper argues that 3D visualization may have a profound effect in information visualization when more advanced display and interactive facilities (e.g. stereo vision) become available [13].

2.2.3 Biochemical Pathway Graph Models

Another important consideration when drawing biochemical pathways as graphs is the structural representation of biochemical networks. Several graph models have been proposed for representing biochemical pathways. These are compound graphs, reaction graphs, bipartite graphs, and object models [8]. Table 2-2 below shows some examples that illustrate the difference between some of these models. A more detailed explanation of each model follows.

Table 2-2: Examples to illustrate the difference between compound graphs, reaction graphs, and bipartite graphs (Paraphrased from [8])

Set #	Reactions	Compound Graph	Reaction Graph	Bipartite Graph
1	R1: A + B → C R3: C → D			
2	R1: A → C R2: B → C R3: C → D			
3	R1: A → B R2: C → D R3: B + D → E			
4	R1: A → B R2: C → B R3: B → E			

Compound graphs are used to model a set of chemical reactions, with nodes representing the chemical compounds and edges represent the reactions [8]. Edges can be directed or

undirected, depending on whether or not the reaction is reversible. An undirected edge between A and B means that both A and B occur as substrates and products in the reaction, whereas a directed edge from A to B means that A is the substrate and B is the product in the reaction [8] (See Table 2-2 for some examples).

Although compound graphs can be used to represent biochemical pathways, they have their own limitations. Compound graphs cannot combine metabolic and regulatory pathways because no distinction is made between nodes that represent compounds from nodes that represent genes and between edges that represent reactions from edges that represent regulations [8]. In addition, compound graphs have limited coverage, as there is no information on enzymes catalyzing reactions [8]. The most important limitation is that compound graphs have an ambiguous representation of reactions because the structure of the reaction is lost, since different sets of reactions can lead to the same compound graph [8]. Table 2-2 illustrates the ambiguity for sets 1 and 2, as both of these sets lead to the same compound graph.

Reaction graphs are similar to compound graphs, except the nodes are the reactions. An edge exists between two reactions if the same compound is a product in one reaction and a substrate in another. Reaction graphs have similar limitations to compound graphs [8]. Table 2-2 illustrates the ambiguity for sets 3 and 4, as both of these sets lead to the same reaction graph.

In bipartite graphs, there are two classes of nodes: one class represents compounds and another represents reactions. Edges can be directed or undirected. As such, bipartite graphs represent reactions without ambiguity [8], as illustrated in Table 2-2. Although bipartite graphs solve issues in compound and reaction graphs, they still have their own limitations. In bipartite graphs, possible controls of reactions, such as catalysis and inhibition, cannot be explicitly represented [8]. Without extensions, bipartite graphs cannot simultaneously model metabolic and regulatory pathways [8].

To deal with the limitations of bipartite graphs, object models are used. Object models are considered a generalization of bipartite graphs, where the nodes are typed to allow more detailed description [8]. Object models define objects and their attributes, as well as hierarchies and relations between the objects [8]. Examples of objects used in object models are compounds, enzymes, reactions, etc. As such, graphs in object models are drawn the same way as bipartite graphs, with the nodes and edges being the actual objects. Object models have been used for the design of biological databases such as MetaCyc [8]. Object models are a good way to represent biochemical pathways because they represent reactions unambiguously and their coverage is large, allowing for representing both metabolic and regulatory pathways [8].

2.3 3D Navigation and Interaction Methods

This section describes the most important 3D navigation and interaction methods that can be used in information visualization interfaces in general. Navigation is a process whereby users determine where they are in relation to their surrounding environment, and how to get to particular objects or places given their location in space [6]. It is important to note here what we mean by 3D navigation. Although for many people, 3D navigation means manipulating a 3D model (including rotation in the third dimension), in some literature, 3D navigation refers to techniques such as zooming, elision, multiple windows, etc, as will be presented shortly. We use the latter meaning in this thesis.

The 3D navigation methods presented below can be used with 2D or 3D graphs, and with other forms of information visualization, such as textual visualization. The navigation and interaction methods must be usable in the sense that they must allow the user to manipulate the software and visualize information with relative ease. We will first present the various 3D navigation and interaction methods available in information visualization, and then we present a discussion on the methods that are mostly appropriate for biochemical pathways.

2.3.1 Zooming Techniques

Zooming is one of the methods that are fairly simple and well suited for displaying less or more detail of the graph. There are two forms that are widely used: Geometric zooming, which enlarges the graph, and semantic zooming, which provides more detail and information when approaching a certain area of the graph [13]. Zooming presents some difficulties when used in interactive environments. One has to deal with the fact that in order to zoom to a distant area of the graph, the user has to zoom out before zooming in, which consumes time [13]. Another problem with zooming is the loss of contextual information when one zooms in [13]. Nevertheless, providing some sort of zooming is desirable and essential when dealing with large graphs.

2.3.2 Focus+Context Techniques

One way to solve the problems associated with zooming is to focus on some detail without losing the context ([13], [34]). An example of a layout that employs these techniques is hyperbolic views [13]. There are 5 techniques that fall under Focus+Context techniques. These techniques are thought to provide usefulness particularly for experienced users [13]. The techniques are summarized below.

2.3.2.1 Distortion

Distortion, or fisheye distortion, uses simple algebraic functions to expand the space around the point of interest and decrease the space given to those objects far away from this point [34]. The pitfall of fisheye distortion is that the edges connecting the nodes will also be distorted, resulting in a general curve and adding new unwanted edge-crossings, since standard graphic systems do not offer the necessary facilities to transform lines into these curves [13].

2.3.2.2 Rapid Zooming

Rapid zooming allows the user to rapidly zoom in and out of points of interest, by changing the camera's position or focal length or by scaling the object [34]. Experiments suggest that the optimal zooming rate should be eight scaling factors per second [38].

2.3.2.3 Elision

Elision is a technique that hides parts of the structure, typically by collapsing sub-graphs into a single node [34]. Elision provides nesting, abstraction, and filtering, which are quite common in software and is useful in representing hierarchies. The purpose of elision is to reduce the number of visible elements being viewed, thus providing clarity and increasing performance of layout [13]. There are two kinds of elision techniques: Structure-based (focuses only on structural information in a graph) and content-based (focuses on the use of semantic data, requiring application-specific data and knowledge) [13].

2.3.2.4 Multiple Windows

The multiple windows technique means having one window to represent an overview of information and several other windows to show the details [34]. They are used to allow for the extension of the visualization of information, either by using alternate views of the same data, or by viewing distinct locations ([2], [38]). It is suggested that the maximum scaling factor between a detailed window and the overview should be of 25 times [38]. When using multiple windows, the user may not clearly see the association between the views in the windows. To deal with this issue, Parker et al. suggests connecting the detailed window to the overview window [34].

Multiple windows add complexity to the interface, but may enhance the user's performance. For instance, if the software allows for overview-and-detail coordination, such as simultaneous updating, highlighting, etc., user performance is improved ([5], [38]). However, to allow for such simultaneous coordination is complex and adds significant overhead to the application. Thus, the designer must carefully weigh out the pros and cons of using multiple windows and make a good justification for his or her choice. The designer must consider the cost-benefit tradeoffs between the cognitive overhead reductions that results from the user of multiple windows and the impact on the system's performance in

terms of time and space [2]. The 4 rules that aid in deciding whether or not to use multiple views are summarized below, and are taken from [2].

Rule of diversity

The rule of diversity states that the designer should "use multiple views when there is a diversity of attributes, models, user profiles, levels of abstraction, or genres". When there is lots of diverse information to represent, using multiple views means that the user has less information to remember.

Rule of Complimentarity

This rules states that the designer should "use multiple views when different views bring out correlations and/or disparities". This allows the user the compare information without having to memorize or switch among the components.

Rule of Decomposition

This rule states that the designer should "partition complex data into multiple views to create manageable chunks and to provide insight into the interaction among different dimensions". This allows the user the compare information without having to memorize or switch among the components.

Rule of Parsimony

This rule states that the designer should "use multiple views minimally". The reason for that is that multiple views add overhead in terms of time and space complexity. Thus, one must be careful and must justify the reasons and costs of additional views. The paper suggests that this rule should in fact take precedence over all other rules.

We now describe the rules involved with how the designer should use multiple views. Again, these are taken from [2].

Rule of Space/Time Resource Optimization

This rule states that the designer should "balance the spatial and temporal costs of presenting multiple views with the spatial and temporal benefits of using the views". The designer must basically keep in mind the space and computational time required to present the multiple views side-by-side, as opposed to representing them sequentially, or using context switching. The paper suggests that the designer should also consider the platform used when making the decision.

Rule of Self-evidence

This rule states that the designer should "use perceptual cues to make relationships among multiple views more apparent to the user". Examples of visual cues are highlighting and coupled interaction. Cues are used to draw the attention of the user to the changes made.

Rule of Consistency

This rules states that the designer should "make the interface for multiple views consistent, and make the states of multiple views consistent". Consistency is very critical in designing user interfaces. In fact, all usability guidelines described in [21] emphasize the importance of consistency, specifically when feedback and response times are concerned. Inconsistent views may increase the learning curve, by making the system harder to understand, and, more drastically, may mislead the user into making the wrong design decisions.

Rule of Attention Management

This rule states that the designer should "use perceptual techniques to focus the user's attention on the right view at the right time". Visual cues such as highlighting and animation can be used. The use of cues speeds up the learning curve and the speed the user can accomplish a task.

Table 2-3 below summarizes the positive and negative impacts of the various guidelines for using multiple windows.

Table 2-3: Positive and negative impacts of multiple window guidelines (Taken from [2])

Rule	Positive Impact	Negative Impact
Diversity	Memory	Learning, computational, space overhead
Complimentarity	Memory, comparison, context switching	Learning, computational, space overhead
Decomposition	Memory, comparison	Learning, computational, space overhead
Parsimony	Learning, computational, space overhead	Memory, comparison, context switching
Space/Time Resource Optimization	Comparison, computational, space overhead	-
Self-Evidence	Learning, comparison	Computational overhead
Consistency	Learning, comparison	Computational overhead
Attention Management	Memory, context Switching	Computational overhead

2.3.2.5 3D Interactive Visualization

These techniques involve changing the 3D properties of the scene to display different parts of the graph, such as changing the focus by changing the viewpoint or the foreground [34].

2.3.3 Impact of Spatial Ability and Cues

There has been evidence suggesting that the user's spatial ability has an impact on the users' performance with 3D interfaces. Spatial ability scores can be determined from a simple multiple-choice questionnaire about the consequences of punching a hole in a paper [6]. Results of an experiment conducted by Chen et al. to evaluate the user interface of a 3D hypertext browser show that users with high spatial ability completed their tasks quicker and explored a larger number of categories than users with lower spatial ability [6]. Thus, the designers of information visualization interfaces must carefully explore their classes of users

and perhaps design interfaces intended for users with lower spatial ability so that they can accommodate the various spatial abilities. In addition, the experiment conducted by Chen et al. showed that the presence of cues, such as borders and gridlines, significantly improved the navigation performance [6].

2.3.4 General Searching and Navigation Considerations

This section is devoted to outlining other behaviors of users, observed while navigating or searching for information.

The results of an experiment conducted by Chen et al. showed that users searching for information by typing keywords are affected by varying the keywords in the search [6]. Sometimes, if the title of an article the user is looking for does not match a keyword, the user ignores the article even if the title and the keyword have similar meanings [6]. The results also showed that when navigating through a graph, users started with the central node as the natural starting point for browsing, and that users tend to ignore the target initially until they gained more experience and confidence with the software [6].

Special attention must be given to the user studies conducted by Neilson. Neilson suggested that when users get to a page containing information, "users look straight at the content and ignore the navigation areas" [32]. This means users tend to focus on the window that displays the information and ignore 3D navigation controls.

From the above results, certain conclusions can be made. The first is that one must be careful in the choice of words, titles, names, etc. that are given to objects or data in information visualization tools. In addition, searching should be comprehensive in a sense that it should allow the user to search for all words the user enters in no particular order. The second is to provide navigation capabilities in the main window where the information is visualized, as opposed to having such navigation controls in a menu or a place that is far from the main window (i.e. far from where the users' eyes are directed).

2.3.5 Application of 3D Navigation and Interaction Methods to Biochemical Pathways

The 3D navigation techniques mentioned above certainly aid in navigating large information networks. With respect to biochemical pathways, several of these techniques are very essential. The first is zooming (See section 2.3.1). Since biochemical pathways are highly interconnected, users may wish to zoom in and out to visualize parts of a pathway or more than one pathway at a time. Thus, some sort of zooming is very desirable in a biochemical pathway visualization tool. In addition, the presence of multiple views is advantageous (See section 2.3.2.4). Biochemical pathways consist of different classes and different types of compounds, with different attributes. In addition, biochemical pathways contain different levels of abstraction. Thus, according to the rules of diversity and decomposition, the presence of multiple views to represent the pathways allows the user to better comprehend the information. However, the user must be given the option of visualizing one view and the use of multiple views should be minimized, keeping in mind the rule of parsimony.

Both zooming and multiple windows are two techniques that are essential in interfaces that involve comparing information that is widely separated in space, such as comparing distant nodes in a graph. Since a very frequent use of biochemical pathway visualization tools would be to compare pathways that may not necessarily be close in space, the use of zooming versus multiple windows must be compared. In particular, the results of an experiment conducted by Ware et al. are useful to determine which technique is best [38].

Ware et al. believe that the results of the study are influenced by the user's working memory. The user's working memory can either be visual, verbal or both, depending on the interface. However, for most information visualization interfaces, it is the visual working memory that counts. Miller has put a limit of 7 ± 2 items on the working memory [38]. This turned out to be more closely related to the verbal memory, as it relates to the phonological length of items [38]. The limit on the visual memory was determined by a sequential comparison experiment conducted by Vogel, where a sample set was displayed to the user, followed by a blank field to clear the iconic memory, and then the probe set [38]. The user was then asked whether or

not probe set matched the sample set [38]. The experiment has revealed that the visual working memory is limited to 3-4 objects at a time and that the arrangement of objects, not the attributes (color, line orientation, shape), affects the limit on the visual memory [38].

The multi-scale comparison test used in the experiment described by Ware et al. is similar to the sequential comparison experiment, with the difference that there may be more than one probe set and that the sample and probe sets are separated by distance rather than by time [38]. Users were asked to compare 6 probe sets of n objects with a sample set and find the one that matches the sample set [38]. Only one probe set exactly matched the sample set, while the rest differed in exactly one object, either in shape, in color, or in both [38].

The results of the experiment showed that zooming is more effective for comparing smaller sets of objects, whereas multiple windows are more effective when an object set is too large to fit in the visual working memory [38]. The main disadvantage to multiple windows is the additional setup time of creating and managing the additional windows. The percentage of errors (user selecting the wrong probe set as a match for the sample set) increased with n, the number of objects, but the percentage was much greater for zooming [38]. This suggests that multiple windows may be more accurate than zooming. From the results of this experiment, we can conclude that the choice between zooming and multiple views depends on the number of objects that one wishes to compare and on whether or not we can afford having the additional overhead created by multiple windows.

In addition to zooming and multiple windows, the use of cues, such as borders and gridlines, is necessary in a biochemical pathway visualization tool, in order to enhance the navigation performance (See section 2.3.3). Moreover, navigation capabilities must be provided close to the area where the pathway is visualized, as opposed to having such navigation controls in an area that is far from the user's focus (See section 2.3.4). Furthermore, biochemical pathways contain a large amount of information and hence the tool should provide a good search engine to allow users to find information quickly and easily. The searching must be comprehensive and less sensitive to the order of words typed (See section 2.3.4).

Chapter 3 User Preferences in a Biochemical Pathways Visualization Tool

This chapter is aimed at capturing users' preferences and requirements for designing a biochemical pathway visualization tool. Three different types of user studies are conducted. These are a videotaped evaluation of existing tools, a questionnaire, and a brainstorming session.

The choice of users depends on the user study conducted. For the videotaped evaluation, only five users are typically needed with a basic knowledge of biochemistry and computers. It has been shown in literature that having five users participate in the videotaped evaluation is sufficient to collect enough data about the usability malfunctions found in the tools [21]. The reason that users only need a basic knowledge of biochemistry is because the visualization tool is targeted towards students for learning purposes.

For the questionnaire, a larger number of users is needed than for the videotaped evaluation in order for statistical analysis to be possible. Users should have divers knowledge and academic and industrial backgrounds, primarily related to science and/or medicine. This allows us, in the early stages of the design, to take into account features that are desired in a biochemical pathway visualization tool, as well as features that would be desired in the future for research purposes. Twelve users were available to complete the questionnaires (five out of these twelve users participated in the videotaped evaluation).

For the brainstorming session, a sufficient number of users from a diverse pool is also needed. Twelve users participated in the brainstorming session, including graduate students, professors and technicians.

Ideally, we would have more users for all of the above user studies in order to gain more confidence in our results. However, additional users were not available and we believe that

the number of users used in this research is sufficient to provide us with enough information to design a usable UI. The subsequent sections discuss each method in more details.

3.1 Videotaped Evaluation of Existing Tools

3.1.1 Features Tested

The videotaped evaluation was performed for five different biochemical pathway visualization tools; four of them (BioCarta, ExPASy, KEGG, and WIT) are web-based, and one (MetaCyc) is a stand-alone application. All of the web-based tools are free and MetaCyc is free for academic purposes.

The main types of functionality that we studied in this evaluation were searching for and navigating biochemical pathways; searching for and finding information about chemical compounds found in the pathways; whether or not the pathway diagrams are linked, and additional features that each tool provides (if any), such as coloring compounds in pathways, highlighting pathways, viewing an overview of pathways, or viewing several pathways simultaneously. We focused on the usability of these features.

3.1.2 Users

Since the tools only provide navigation capabilities for biological compounds and pathways, without any analysis of this information, users only need basic knowledge of biochemistry that could be obtained from an introductory biochemistry course.

Prior to finalizing the set of instructions used in the videotaped evaluation, we performed a pilot study with one user, who has a background in computer science and who has worked extensively with biochemical pathways. The purpose of the pilot study was to test the camcorder and to make sure that the session takes an hour to an hour and a half to complete.

From the pilot study, the set of instructions was reduced from 40 instructions to 30 (See Appendix B.3).

The videotaped evaluation of the tools was performed after the pilot study. Five users, with no previous knowledge or experience with these tools, participated in this activity. These users were representatives of the target end-users since they had a good knowledge of biochemistry; they also had basic knowledge about how to use computers and the Internet.

Four of the users had finished their third year in biochemistry or biology and had taken an introduction to biochemistry course. One user is a biochemistry lab technician and has experience with biochemical pathways

3.1.3 Procedures

The videotaped evaluation was performed using the instructor's laptop (a TOSHIBA Pentium 4 CPU 1.60GHz). Prior to the evaluation period, the users were informed of the purpose of the study and were asked to read and sign the informed consent forms (See Appendix B.1). The users were then given a quick PowerPoint tutorial on the five tools (See Appendix B.2). The tutorial consisted of a description of the main focus of each tool, whether the tool is web-based or stand-alone, how to use the tool to perform simple pathway or compound searches, whether or not the tool allows editing or adding pathways, whether the tool visualizes pathways statically or dynamically, example(s) of the way the tool represents metabolic and/or regulatory pathways, and a description of the pathway visualization each tool provides and which items are clickable.

The evaluation session took place in a private office at the University of Ottawa. Each session lasted from one hour to one hour and a half. A total of 30 tasks were given to the user. To minimize interference due to learning, the order of the tasks was changed from one user to the next. A complete list of the tasks is found in Appendix B.3.

3.1.4 Results

The videotaped sessions lasted about 6 hours and each hour of the videotape took about 10 hours to evaluate, for a total of 60 hours. For each user, the total time to accomplish each task was measured in seconds and the total number of malfunctions uncovered in each tool was measured.

3.1.4.1 Task Completion Time Results

In order to avoid bias, the time it took to load a page and the time that involved conversation with the user were subtracted from the total time to accomplish the task. Please refer to Table B-1 of Appendix B.4, which lists the speed and average of task completion time in seconds for each user using the tools. Please note that all entries in red italics in the table are not valid and were not taken into consideration. N/V (Not Valid) means that the measurement is not valid because considerable help was given to the user or because the user did not accomplish the task correctly. Infinite times are times for tasks that took excessive time (greater than 180 seconds for most tasks) so the user gave up doing them. Other numbers in red italics are not valid because either the user did not really accomplish the task or only accomplished part of the task.

Figure 3-1 below shows the average task completion times in seconds for each tool. Tasks that took 400 seconds are these for which all users spent excessive time and eventually gave up. These tasks did not actually take 400 seconds to complete. However, in order to show that they took a long time and were not accomplished, they are shown in the figure as such. Out of these tasks, the ones that are interesting to look at are Tasks 1 to 7 because they allow one to compare the completion times of the different tools. Tasks 2 and 6 do not show data on MetaCyc, not for the reason that MetaCyc does not support them, but rather for the reason that these tasks were accidentally omitted for MetaCyc. Task 4 is not supported in MetaCyc and WIT, and Task 7 is not supported in BioCarta and ExPASy. Tasks 8 to 11 represent tasks that are only available in a certain tool. The last of these, Task 11, is present in KEGG and it took an infinite time according to the figure. Only 2 users were actually capable of completing the task, without any help. The other 2 users gave up doing this task and the last

user was given many hints to accomplish it, which invalidated the result from this user. In fact, this task uncovered a considerable number of malfunctions (See Figure 3-2).

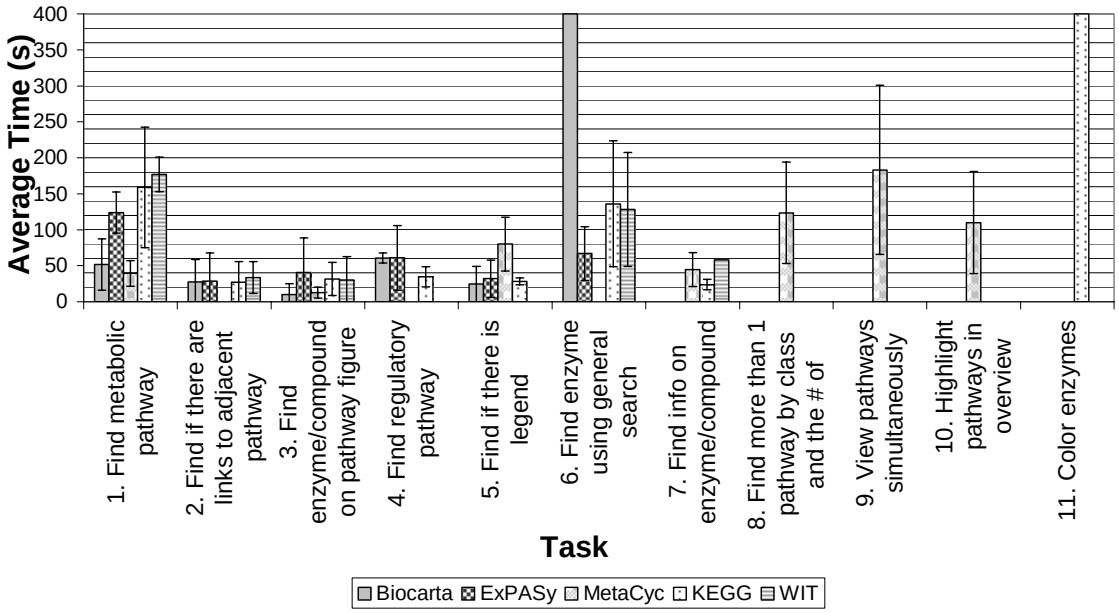


Figure 3-1: Comparing average task completion times for 5 users using 5 existing tools. The error bars represent the standard deviation

Since we are dealing with averages, we would like to make inferences about the differences among these averages [29]. For instance, looking at Figure 3-1 above, can we conclude that for Task 1, users of BioCarta can finish this task in 73 seconds less time than users of ExPASy, meaning that BioCarta is better than ExPASy for this task?

Since we are dealing with more than two population means, we need to first conduct an analysis of variance (ANOVA) test to determine whether there are significant differences between the means of the populations, or whether the differences were purely due to random chance. Please see Appendix D.1 for a more detailed description of the ANOVA test.

We applied the ANOVA test to all tools and for each task. We used a value of α=0.05, meaning a 95% confidence level. The results are summarized in Table B-3 of Appendix B.4. From the results in the table, we can conclude that for tasks 1 and 5, there is a significant

difference between the means of the populations. For the rest of the tasks, we conclude that the observed differences between the sample means are not significant.

In order to draw conclusions about which tools are better for tasks 1 and 5, we would like to conduct a T-Test between each pair of tools. Please see Appendix D.2 for a more detailed description of the T-Test.

We applied the T-Test to each pair of tools and for each task. We used a value of $\alpha = 0.05$, meaning a 95% confidence level. The results are summarized in Table B-2 of Appendix B.4. From the results in the table, we can conclude that for Task 1 (finding a metabolic pathway), we are 95% confident that if we take a random user from the population and ask the user to accomplish the task using the five tools, the user will accomplish the task faster using BioCarta ($\bar{x}$ = 51.40 s) and MetaCyc ($\bar{x}$ = 39.40 s) than using ExPASy ($\bar{x}$ = 123.75 s), KEGG ($\bar{x}$ = 159.00 s), or WIT ($\bar{x}$ = 177.00 s) and faster using ExPASy ($\bar{x}$ =123.75 s) than using WIT ($\bar{x}$ =177.00 s). For Task 5 (finding if there is a legend), a random user from the population will accomplish the task faster using BioCarta ($\bar{x}$ = 24.33 s), ExPASy ($\bar{x}$ = 32.00 s), and KEGG ($\bar{x}$ = 28.20 s) than using MetaCyc ($\bar{x}$ = 80.00 s). T-Test results for tasks 4 and 7 are ignored because results of the ANOVA test show that the differences between the means for these tasks are not significant.

3.1.4.2 Analysis of User Interface Malfunctions

A malfunction is a usability defect, i.e. an obstacle to the smooth operation of the user/computer system [22]. Any incorrect behavior of the system, whether it is elated to 'utility' of 'usability', is recorded as a malfunction. Examples of malfunctions related to utility include when the system crashes while accomplishing a certain task, or when a feature is missing entirely. An example of a malfunction related to usability is when the user expresses frustration or confusion when accomplishing a certain task. This confusion results from having an incorrect mental model. The reason could be a misleading menu item, a misleading button, etc. The purpose for following UI guidelines (see [21]) is to minimize such malfunctions. Figure 3-2 below shows the number of malfunctions uncovered by the videotaped evaluation of each tool.

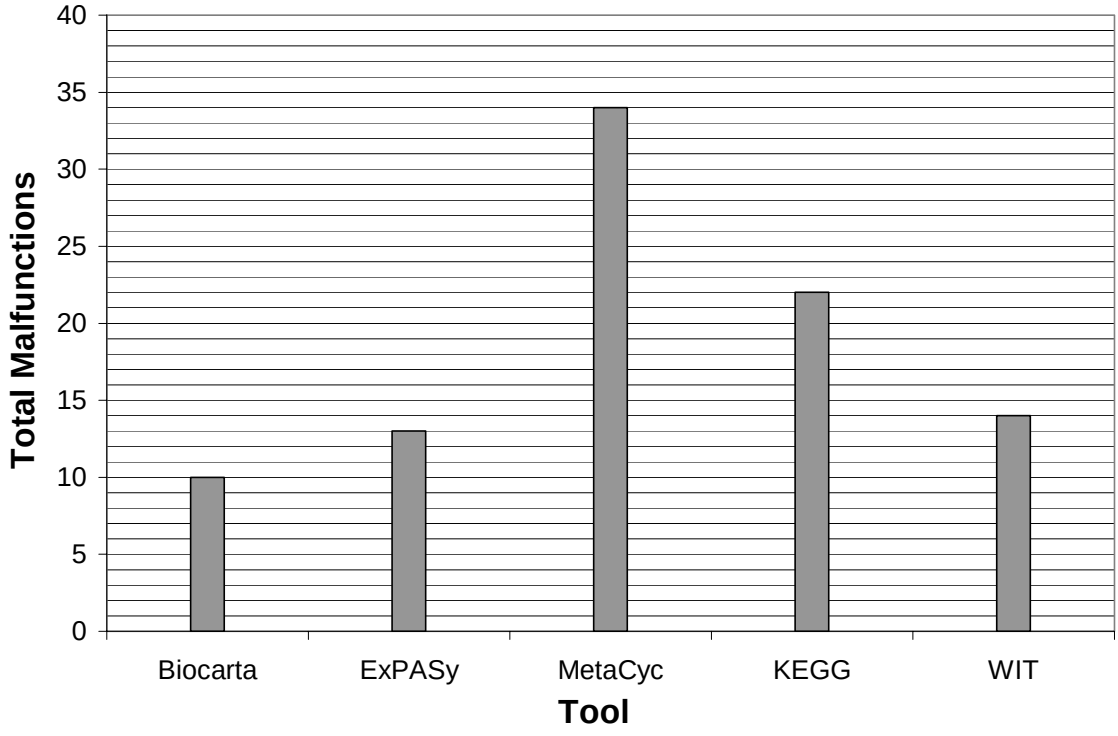


Figure 3-2: Comparing total malfunctions found in existing tools (Data collected from videotaped evaluation with 5 users)

For each malfunction, the category the malfunction belongs to, the task the malfunction occurred at, the user #, the problem description, the UI principles violated, the severity level, and recommendations for change were determined. Please refer to Appendix B.5 for a complete list of all malfunctions for each tool and their detailed description.

Figure 3-3 below compares the total malfunctions uncovered in each task in the five tools. For example, for finding a metabolic pathway, WIT had the most number of malfunctions, indicating more difficulties in comparison with other tools. MetaCyc, on the other hand, had the lowest number of malfunctions. When comparing this figure with Figure 3-2, we see that although Figure 3-2 shows that MetaCyc had the most number of malfunctions (34 malfunctions), most of these malfunctions (23 malfunctions) are related to the additional functionalities that MetaCyc provides, and not to the basic pathway/enzyme searches. Similarly, for KEGG, 9 out of 22 malfunctions are related to coloring enzymes, a

functionality that only KEGG provides. Thus, in order to compare the number of malfunctions in the five different tools, we must only consider the tasks that are common between the tools (Tasks 1 to 7 in Figure 3-3 below). Note that when determining the malfunctions by task, each malfunction is counted only once. Hence, if a malfunction is found in Task 1, it is counted for Task 1. If that same malfunction is found in subsequent tasks, it is not counted for these tasks.

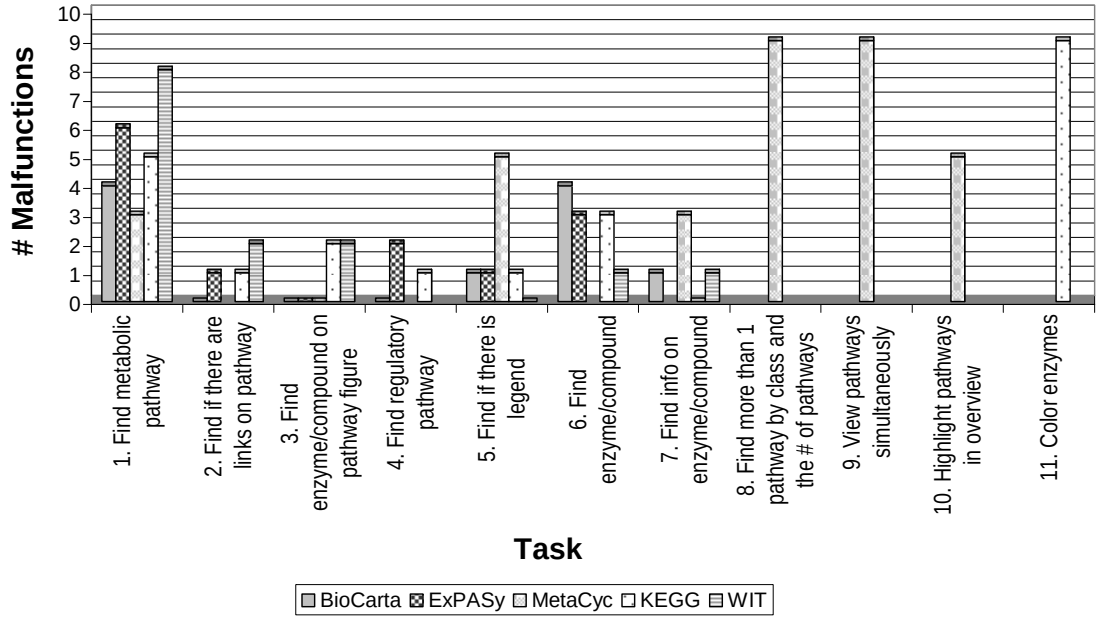


Figure 3-3: Comparing malfunction distribution according to task in existing tools (Data collected from videotaped evaluation with 5 users)

For finding metabolic pathways, WIT had the greatest number of malfunctions and MetaCyc had the fewest malfunctions. WIT had the problem of providing non-meaningful search result links, with no description of what they mean, which users found ambiguous and confusing. WIT and ExPASy do not allow one to search for pathways by name, which users would have found desirable. Most tools, such as BioCarta, ExPASy, MetaCyc, and KEGG, did not offer searching flexibility (i.e. searching for pathways containing all given keywords), which users found annoying.

For finding if there are links on the pathway figure, WIT had the most malfunctions as compared with the other tools. BioCarta, on the other hand, had no malfunctions. This is because BioCarta does not have adjacent links on the figure and this was obvious to the user. KEGG and ExPASy had crowded figures and it was not very easy to determine where the links are. WIT had clickable compounds, which made the user think that these may lead into linked pathways, but the user did not want to click on the links because the tool was taking a long time. Because of an oversight of the researcher, this task was not done for MetaCyc, because of improper experimental design.

To find an enzyme/compound on the pathway figure, WIT and KEGG had the most malfunctions. BioCarta, ExPASy and MetaCyc had no malfunctions. The reason for that is that WIT and KEGG show enzymes as Enzyme Classification (E.C.) numbers without the name, which made it harder for the user to find the enzyme on the figure, whereas the other tools used the enzyme name.

For finding a regulatory pathway, which MetaCyc and WIT, do not support, ExPASy had the most malfunctions, followed by KEGG, followed by BioCarta, which had no malfunctions. The reason is that ExPASy does not allow searching for a pathway by name. KEGG, on the other hand, had some malfunctions that were mostly due to ambiguous labels and inappropriate location of search links.

For finding if there is a legend on the pathway figure, MetaCyc had the most malfunctions. This is because the link that allows one to search for a legend is titled “Show Key”, which does not indicate the word “legend”. All of the tools, except WIT, which does not have a legend, had the problem that the legend is not placed on the pathway figure itself, which all users found unintuitive.

For finding compounds or enzymes using general search, BioCarta had the most malfunctions, followed by ExPASy and KEGG, followed by WIT. BioCarta’s main problem is that the field that allows one to search for compounds or enzymes is titled “Gene Name”, which users found misleading. KEGG’s main problem is the title and location of the search

links. ExPASy did not provide searching flexibility. WIT had the problem of providing non-meaningful search result links, with no description of what they mean, which users found ambiguous and confusing.

For finding information on enzymes using the enzymes page, MetaCyc had the most malfunctions, mainly due to the use of many bright colors on white background and the use of small fonts. The problems with BioCarta and WIT have to do with the summary pages themselves. BioCarta's enzyme information page only contains links to other pages, with no description of what kind of information these links contain, which is confusing to the user. WIT's page is crowded, not nicely presented, and uses a lot of abbreviations.

Another important criterion that one should consider when comparing the number of malfunctions in the five tools is severity. When assessing the severity level of malfunctions related to usability, five severity levels are used. These are summarized in Table 3-1 below.

Table 3-1: Description of 5 severity levels used to assess malfunctions related to usability issues (Paraphrased from [39])

Severity Level	**Description**
Level 1 (Catastrophic error)	Malfunctions in this level result in loss of data, terrible performance, or damage to the hardware or software, preventing people from doing their work.
Level 2 (Severe problem)	Malfunctions in this level may cause loss of data. There is no workaround to the problem and the system's performance is very poor.
Level 3 (Moderate problem)	Malfunctions in this level do not cause permanent loss of data, but result in wasting time, as a result of internal inconsistencies or features not working as expected. There is generally a workaround to the problem.
Level 4 (Minor but irritating problem)	Malfunctions in this level slow users down slightly, due to minimal violations of usability guidelines that are related to appearance or perception.
Level 5 (Minimal error)	Malfunctions in this level are minor cosmetic or consistency issues that do not cause significant loss of time.

Figure 3-4 below shows the distribution of the malfunctions found in the five tools according to the severity levels described in Table 3-1. As we can see in the figure, most of the malfunctions are in Level 4 (minor but irritating problems). The rest are mostly in Levels 3 (moderate error) or 5 (minimal error). Very few problems are in Level 2, although there are some in MetaCyc and WIT. The severe problems in MetaCyc are due to the fact that the system crashed while attempting to accomplish certain tasks. The severe problem in WIT occurred because the website was very slow, which the users found very annoying.

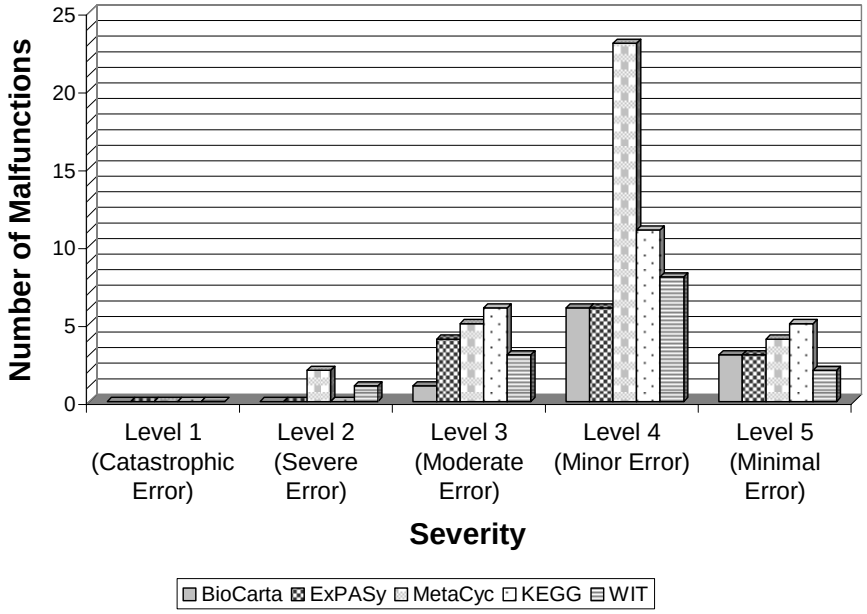


Figure 3-4: Malfunction distribution according to severity in existing tools (Data collected from videotaped evaluation with 5 users)

Finally, although stated in literature [21], Figure 3-5 below reinforces the fact that having five users participate in the videotaped evaluation is enough to find the most number of malfunctions. As the figure shows, User 1 found most malfunctions. The number of additional malfunctions found by each additional user decreases, reaching zero additional malfunctions as we reach User 5.

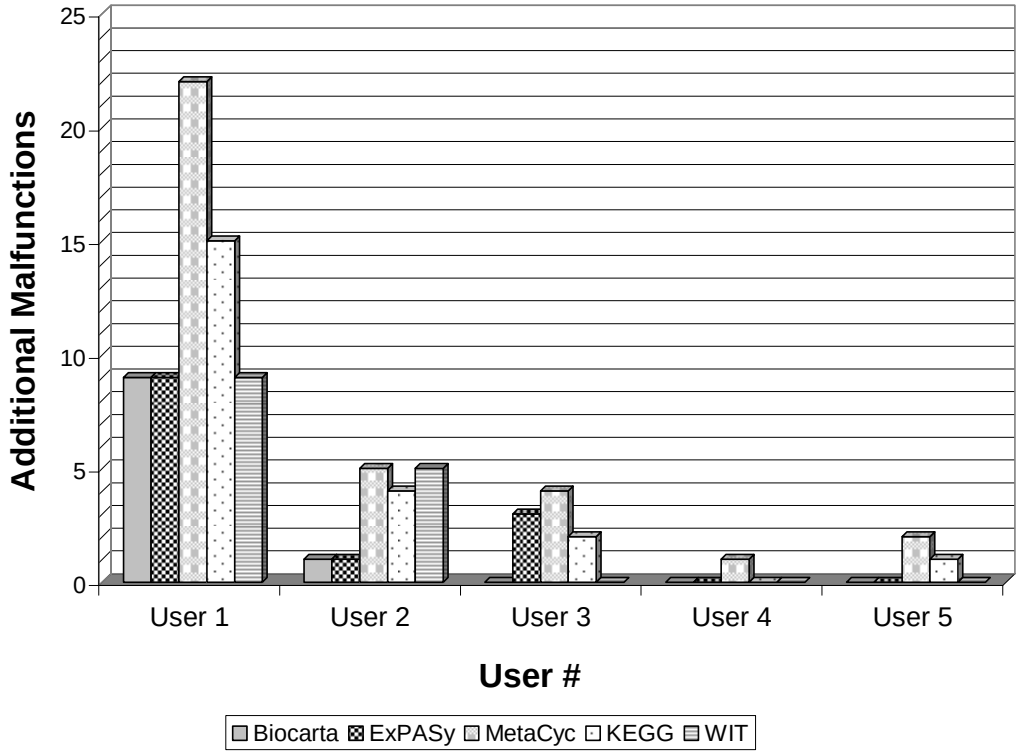


Figure 3-5: Additional number of malfunctions found by each additional user in existing tools

3.2 Biochemical Pathway Visualization Questionnaire

3.2.1 Description

In order to gather more data on user preferences, several users were asked to fill out a questionnaire about what they would like to have in a biochemical pathway visualization tool and what notations and functionality they prefer (See Appendix C).

The questionnaire consists of three sections. The first describes users' usage data, which consists of how often the users use biochemical pathways, their familiarity with existing tools, and the problems they face when using biochemical pathways and existing tools. The second section is used to collect user preference data, in terms of features they think are

important to have and the reasons they would like to have them, as well as tools and notations they prefer and would like to have. The last section consists of personal data about users, such as their age, profession, education, background, and experience with biochemical pathways and computers.

3.2.2 Results

Twelve users were asked to fill out the questionnaire. Out of the twelve users, four are professors (three teach biochemistry and one teaches veterinary medicine), one is specialized in veterinary medicine with a Ph.D. degree, one is specialized in veterinary medicine with a master's degree, one is a biology Ph.D. student, four are 4th year undergraduate students (three are studying biochemistry and one is studying biology), and one is a biochemistry lab technician.

One user uses biochemical pathways daily, one uses them every second day, one uses them twice or three times a week, four use them weekly, one uses them monthly, two use them whenever required for school, one uses them only for teaching purposes, and one never uses them. The undergraduate students mostly use biochemical pathways to study for their courses and for their biochemistry labs, while the professors use them for research and teaching. The specialists use them for research and understanding or interpreting biochemical and physiological data. The undergraduate students stated that they mostly consult books when they need to use biochemical pathways, while the graduate student consults both books and tools. Most professors (three out of four) stated that they mostly consult books, while one stated that he or she consults tools. The specialists stated that they consult both books and tools and the technician stated that he or she consults tools.

Figure 3-6 below shows the average familiarity of the twelve users with the five biochemical pathway visualization tools. Users were asked to rate their familiarity of each tool using six discrete values, which are "Very High", "High", "Medium", "Low", "Very Low", and "Never Used". Then, data was collected from all users and the six discrete values were

converted to numbers from 1 to 6, where 1 is “Very High” and 6 is “Never Used”. The average familiarity is calculated using the following formula:

$$\textit{Average Familiarity} = \frac{(\#\,\text{Very High}) * 1 + (\#\,\text{High}) * 2 + (\#\,\text{Medium}) * 3 + (\#\,\text{Low}) * 4 + (\#\,\text{Very Low}) * 5 + (\#\,\text{Never Used}) * 6}{12}$$

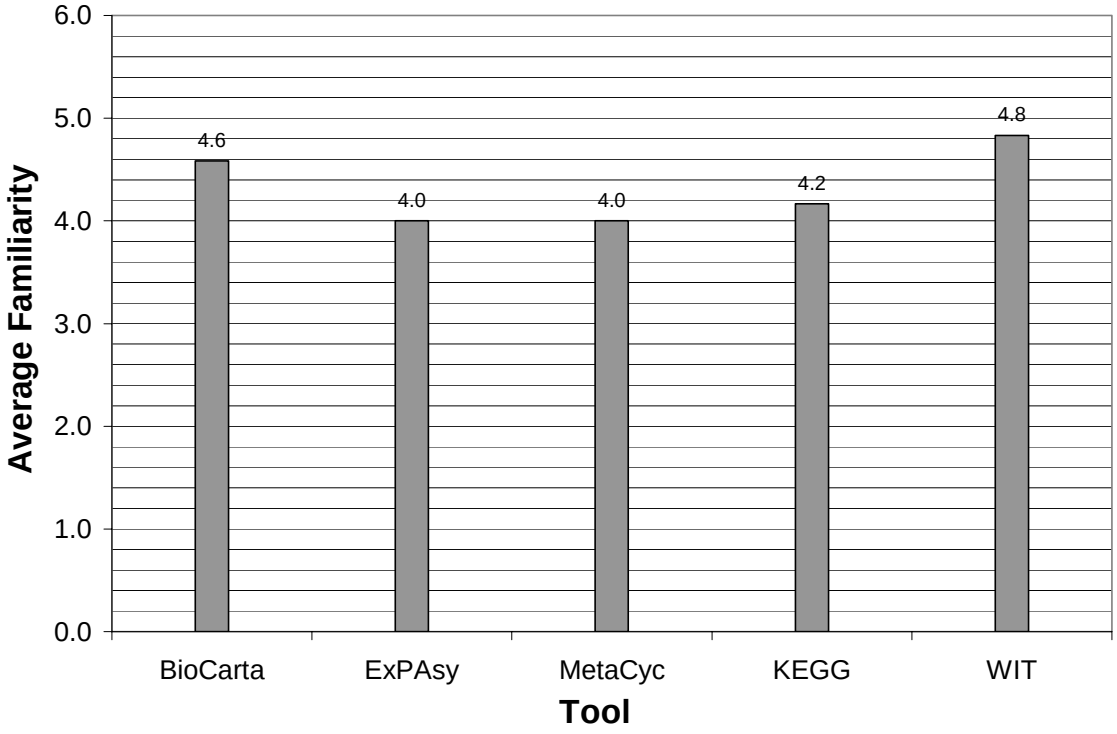


Figure 3-6: Average familiarity with biochemical pathway visualization tools on scale from 1 to 6, where 1 is "Very High" and 6 is "Never Used" (Data collected from 12 users)

Figure 3-6 shows that the average familiarity of all tools is close to 4 (Low) or 5 (Very Low). However, users tend to be mostly familiar with MetaCyc and ExPASy, followed by KEGG, followed by BioCarta, followed by WIT.

The questionnaire asked the users what kinds of problems each tool presents and according to the responses, additional malfunctions are found. Three additional malfunctions of severity Level 4 were found in BioCarta, two additional malfunctions of severity Level 4 and

Level 5 were found in ExPASy, two additional malfunctions of severity Level 4 were found in KEGG, and one additional malfunction of severity Level 5 was found in MetaCyc. Please refer to Appendix C.1 for a complete list of all malfunctions found by the questionnaire and their detailed description.

Users were asked two related questions that rank their preference for the five tools. These two questions were worded differently in order to detect any inconsistencies between the results. The first question asks the users to cast a vote for their most preferred tool for biochemical pathway visualization and the second question asks the users to rank the five tools using five discrete values from "Most favorite" to "Least favorite".

When asked which tool is the preferred tool for biochemical pathway visualization, most users preferred KEGG, followed by MetaCyc, then BioCarta, then ExPASy, and finally WIT, as shown in Figure 3-7 below. Please note that some votes have fractions because some users cast equal votes for several tools. Perhaps the reason for WIT being the least preferred is the fact that the website is slow.

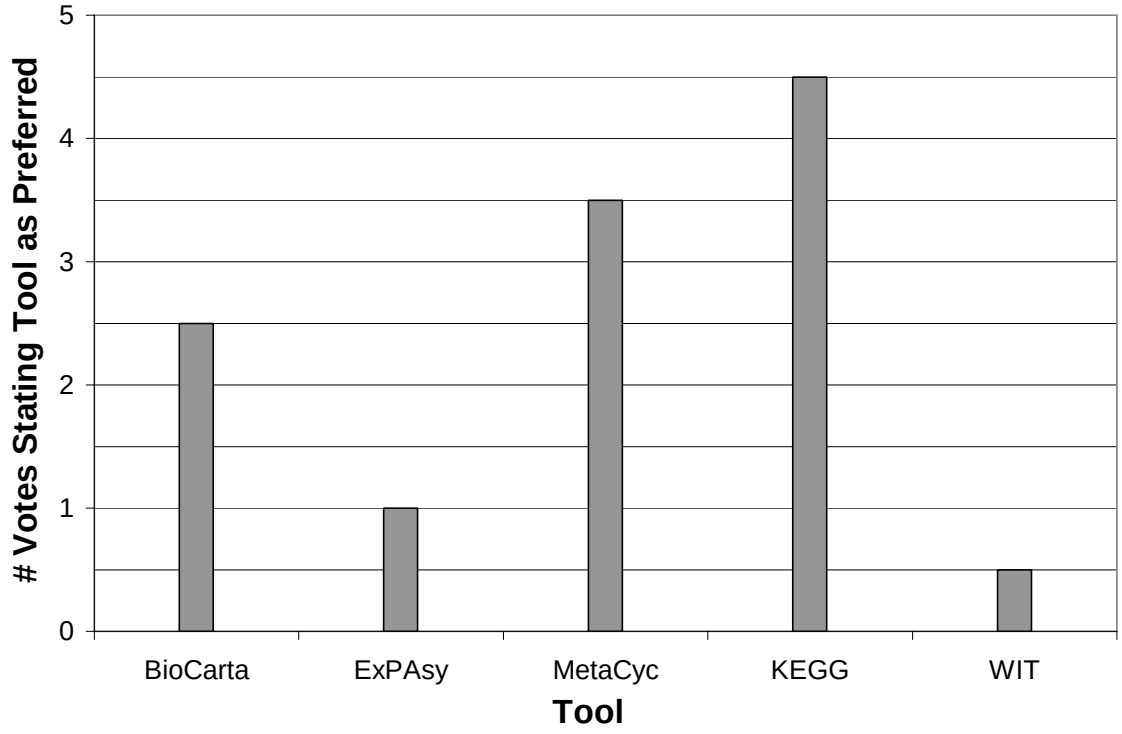


Figure 3-7: Number of votes ranking each of existing biochemical pathway visualization tool as their preferred tool

Figure 3-8 shows the average rank of biochemical pathway visualization tools, where Rank 1 is the most favorite and Rank 5 is the least favorite. The average rank is calculated using the following formula:

$$Average\ Rank = \frac{(\#1) * 1 + (\#2) * 2 + (\#3) * 3 + (\#4) * 4 + (\#5) * 5}{12}$$

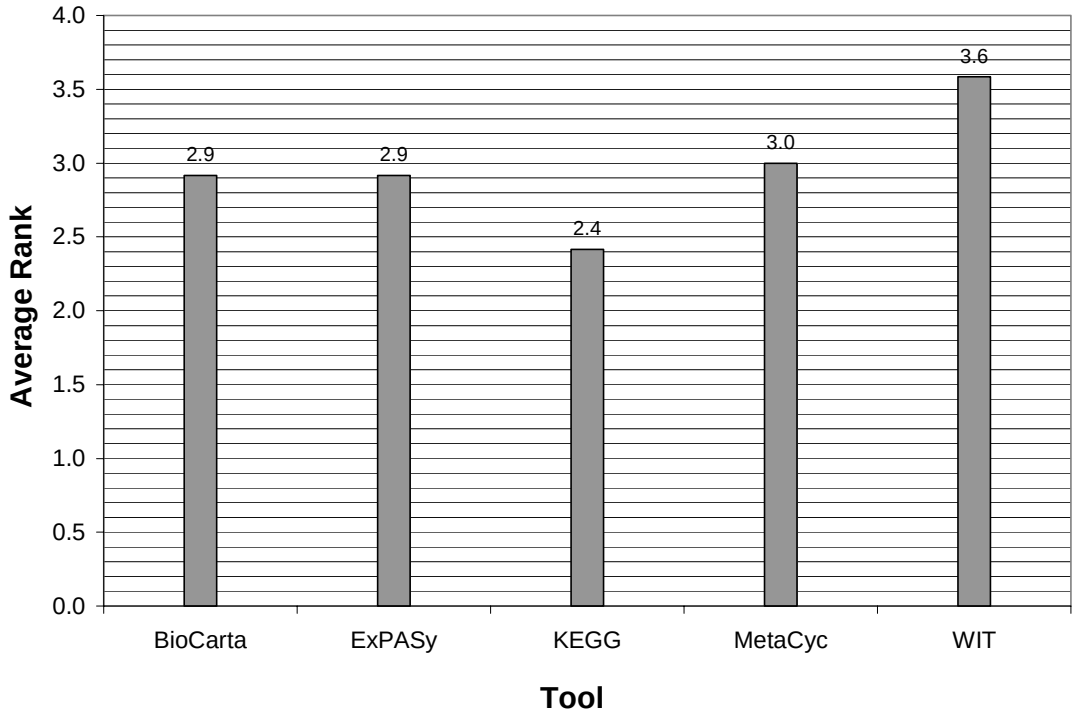


Figure 3-8: Average rank of existing biochemical pathway visualization tools, where 1 is most favorite and 5 is least favorite (Data collected from 12 users)

As the figure shows, KEGG has the lowest average rank, meaning that it's the most favorite tool, followed by ExPASy and BioCarta, followed by MetaCyc, and lastly by WIT.

When comparing Figure 3-7 and Figure 3-8, we can clearly see that KEGG is the most preferred tool, as it received the highest number of votes and the lowest average rank (i.e. the most favorite tool). WIT, on the other hand, is the least preferred tool, as it received the lowest number of votes and the highest average rank (i.e. the least favorite tool). There is no consistency with the two figures with respect to BioCarta, ExPASy, and MetaCyc. Hence, we can only say that preferences for these tools lie in between KEGG and WIT.

Users seemed to prefer KEGG because it is complete and has extensive data content. Users liked the diagrams, even though not much coloring is used, because they had links to adjacent pathways and to literature and other databases. However, some users commented that they did not like the E.C. (Enzyme Commission) numbering system used to classify enzymes, the lack of co-substrate and co-product information on the diagrams, and the fact that there is no legend. Some users commented that they liked ExPASy because it is the most familiar tool that provides links to adjacent pathways and uses nice color-coding. Others did not like it because the diagrams are crowded and navigation is difficult. Some users liked BioCarta because the tool is easy to use, a legend is provided and the diagrams are colorful and easy to read. Others did not like the fact that it lacks regulators (activators and inhibitors) information and that it does not contain all pathways. Some users liked MetaCyc because it is well organized, easy to use, and has nice options. However, other users commented that there is too much cognitive effort involved with using MetaCyc and that it is difficult to use. Most users did not like WIT because the website is too slow and because it is incomplete.

In addition, users were asked to state in general terms, what constraints they face when visualizing/editing biochemical pathways. Their responses are summarized in Table 3-2 below.

Table 3-2: General constraints when visualizing or editing biochemical pathways

C#	User #	Constraint Description
C1	User 1	Lack of a legend on pathway figures.
C2	User 1	Lack of help or tutorial in existing tools.

C#	User #	Constraint Description
C3	User 2	Existing tools have pathway figures that are either too simple in their content (BioCarta) or too complicated (ExPASy).
C4	User 2	Existing tools are not user friendly.
C5	User 3	With some tools, it takes a long time to figure out how to use some features.
C6	User 4	Time.
C7	User 4	Clarity.
C8	User 6	Difficulty to get a global view including interactions between pathways.
C9	User 6, 9, 10	Ability to navigate pathways and easily zoom in and out.
C10	User 7	Ability to visualize and interpret cross-pathway interactions.
C11	User 8	Information is not very well organized.
C12	User 11	Ability to find information on the cellular, tissue, and organism level, as well as hormone control.
C13	User 12	Clear arrangement and visualization of pathways, such that pathway figures are not over-crowded.
C14	User 12	Ability to quickly find information.

These responses show that users want a tool that is clear, easy to use, and allows them to accomplish tasks and quickly search for information. Users would like to have a legend and tutorial or help on how to use the tool when facing difficulties. They would also like to be able to zoom in and out, display links to adjacent pathways, display hormone control, display pathways in different cells, tissues, and organisms, and have the option of displaying more or less detail of pathways.

Figure 3-9 shows the average importance of the features that could be available in biochemical pathway visualization and editing tools. Users were asked to rate the importance of each feature using five discrete values, which are “Very important”, “Important”, “Moderately important”, “Unimportant”, and “Very unimportant”. Then, data was collected from all users and the five discrete values were converted to numbers from 1 to 5, where 1 is “Very unimportant” and 5 is “Very important”. The average importance is calculated using the following formula:

$$\textit{Average Importance} = \frac{(\#\,\text{V. Unimportant}) * 1 + (\#\,\text{Unimportant}) * 2 + (\#\,\text{Mod Important}) * 3 + (\#\,\text{Important}) * 4 + (\#\,\text{V. Important}) * 5}{12}$$

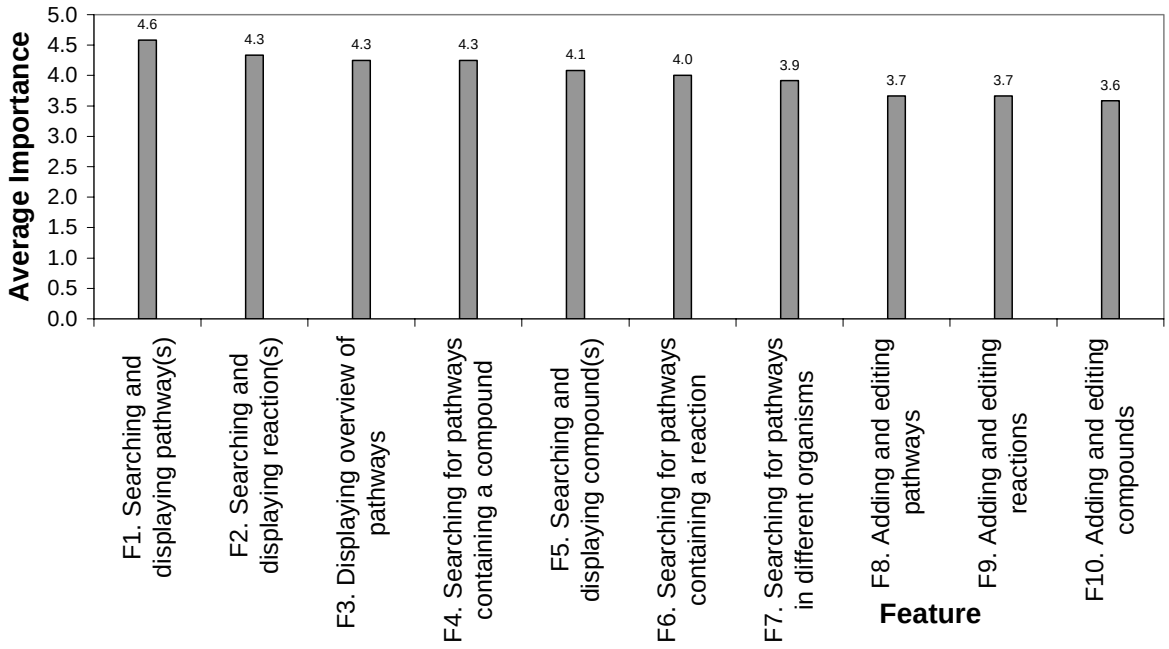


Figure 3-9: Average importance of features in biochemical pathway visualization tools on scale from 1 to 5, where 1 is "Very unimportant" and 5 is "Very important" (Data collected from 12 users)

The average importance of searching and displaying pathways is close to “Very important”. Users commented that, in their view, this is the main purpose of the tool. All the rest of the features have an average importance value, which is close to “Important”. Users commented that the reason is that such features provide basic functionality that is desirable in the tool and allow one to see pathway interactions. In particular, all the different types of searching features are necessary to find information and for navigation. Specialists and researchers commented that adding and editing information is necessary because the biological and biochemical knowledge “is very dynamic and thus adding and editing is a must” in order to “incorporate working hypothesis” and to “evaluate new scenarios”.

Table 3-3 below summarizes other features users would like to have in a biochemical pathway visualization tool, the reason the feature is important, and the average ranking of the importance of the feature.

Table 3-3: Other desirable features in biochemical pathway visualization tool, reason for importance, # of users mentioning the importance of the feature, and average importance of the feature, where 1 is "Very unimportant" and 5 is "Very important"

#	Feature	Reason	# Users	Avg. Imp.
F11	Accessible help menu	Many of the tools have their help menu hidden.	1	5.0
F12	An interactive tutorial	To help students understand how the program works.	1	4.0
F13	Legend presented on all diagrams	Makes understanding the pathway diagram easier.	1	5.0
F14	Glossary	If technical terms are being used, a glossary would be good for understanding.	1	3.0
F15	Color coding (with the legend)	To allow the user to know different compounds and pathway types at a glance.	2	3.5
F16	Searching for enzymes	-	1	5.0
F17	Links to other pathways	To help gain a complete picture of everything (e.g. pathways that affect the reaction under study).	2	4.5
F18	Clickable links to 3D display	To view ligand and enzyme complexes.	1	5.0
F19	Clickable popup menu for reactants	Need to be able to click to get this info (e.g. properties, associated diseases, etc).	1	5.0
F20	Print pathway or parts of pathway	We need this feature for teaching purposes (Not as important for research).	1	5.0
F21	Clickable popup menu for each enzyme	To quickly retrieve the expected information (3D structure, mutations, diseases, sequence alignment).	1	5.0

#	Feature	Reason	# Users	Avg. Imp.
F22	Pathway organization	-	1	5.0
F23	Zooming	Pathway diagrams are huge. Thus, it is essential to be able to zoom in to view the details.	2	5.0
F24	Rolling back into a previous state	The user should be able to undo an action.	2	4.0
F25	Clear description of commands	In order for the user should to know which commands are appropriate.	2	4.0
F26	Self-explanatory	To allow for better usability.	2	4.0
F27	Precise sources	The sources should be credible.	1	5.0
F28	Including reactions and cell structure	-	1	4.0
F29	Having one's own database	To collect all the information needed for personal use.	1	4.0
F30	Including isoenzymes	-	1	3.0

Lastly, when asked about preferred notations in biochemical pathway visualization tools, the results shown in Table 3-4 were obtained.

Table 3-4: Preferred notations in biochemical pathway visualization tool

Notation	Preferences
Color Convention	The color convention of BioCarta received 5.3 votes, ExPASy and KEGG each received 2.3 votes, and MetaCyc and WIT each received 1 vote. The reason for the fraction is that some users gave more than one equal votes.
Intermediate substrate representation	Displaying compounds in their chemical form received 2.5 votes, displaying compounds in textual format received 2.0 votes, and having the option of switching between chemical and textual format received 7.5 votes. None of the users liked having oval boxes or just textual format to represent compounds.

Notation	Preferences
Initial substrate and product representation	Five users liked to have them represented the same way as intermediate substrates and seven preferred to have them represented differently.
Enzyme representation	Seven users preferred having the name on top of the reaction arrow and two users preferred having the E.C. number on top of the reaction arrow. One user preferred having a number on top of the arrow with a legend at the. Two users preferred to have a combination of name and E.C. number. None of the users liked the option of not showing the enzyme at all.
Activator/Inhibitor representation	Five users preferred having a '+' and '-' besides the reaction arrows to represent inhibitors and activators. Four users preferred to have them in textual format. One user preferred a combination of "+/-" and text, one user did not want to see the option on the figure, and one preferred to have the option of removing or displaying activators/inhibitors (as '+/-' and text).
Pathway linkage representation	Having linked pathway names right on the figure received 9.5 votes, whereas having arrows outside of the figure that takes you to linked figures received 1.5 votes. One user preferred having arrows that take you to linked pathways. One user commented that the user would like to see the linked pathway in a separate window.

3.3 Brainstorming Session

3.3.1 Description

The third technique we used to gather data about user preferences was to organize a brainstorming session of about 10-15 professors and graduate students who are interested in biochemical pathway visualization tools. The technique used in the brainstorming session is called the nominal group technique, which was developed by Andre L. Delbecq and Andrew Van de Ven in 1968 [7].

Using the nominal group technique, an individual called the moderator leads the session. Users are arranged in a cycle around a table [23]. Each user is asked to first sign an informed consent form, which states that the user agrees to participate in the session. Then, users are given a pile of blank paper. The moderator starts the session with a general trigger question, where participants are asked to think of as many answers to as possible [23]. As soon as the user thinks of a possible answer, the user must write it on a piece of black paper and pass it to the person on the left. Answers from previous users in the cycle may stimulate new ideas [23]. The process continues until all participants agree that they have no more new ideas. This results in a large informal list of concepts and statements, resulting from users' parallel efforts [23]. Then, a discussion of this list of concepts is accomplished and each member is asked to rank the priorities (in private) [7]. The objectives of the nominal group technique are to allow all members of the session to participate and to incorporate mathematical voting techniques in order to reduce errors in combining individual judgments into group decisions [7].

Twelve users attended the brainstorming session: Five professors (four biochemistry professors and one chemical engineering professor), six biochemistry graduate students, and one user from the Institute für Zell- und Organ Simulation (IfCOS) [14], a biotech company based in Germany that is supporting this research. Dr. Lethbridge was the moderator of the session. The session lasted about 2 hours.

The initial trigger question of the session was "What features would you like to see in a computerized tool for working with biochemical pathways". Users continued passing ideas around the table, until they had no more new ideas. The moderator then asked the users to state which of the ideas they see in the pile of paper in front of them are the most important. These ideas were written on a white board. Then, users were asked to cast 5 votes for what they think are the most important features.

The second general question asked was: "What tasks do you find most difficult when using biochemical pathways – these are the tasks that a tool might be able to support". The answers were also written on the white board. The brainstorming session then consisted of a series of

short and specific questions about searching (both quick and detailed) and about pathway representation, classification, and display. The results are summarized in the next section.

3.3.2 Results

The first trigger question resulted in 15 important features. These features, along with the total number of votes for each feature, are summarized in Table 3-5 below.

Table 3-5: A list of most important features, found by 12 users in the brainstorming session, along with total number of votes for each feature

Feature #	Feature	Votes
F31	Visualize dynamic relationships (e.g. between proteins to help understand regulatory control).	9.0
F32	Easy to use.	7.0
F33	Provide overview of biochemical pathways • To see the big picture – can get into more detail, perhaps with "highlight". • To understand integration/links between pathways.	6.5
F34	Provide good search engine.	5.0
F35	Comprehensive.	5.0
F36	Captivate the audience (i.e. interesting, exciting, and informative) for teaching purposes.	4.5
F37	Interactive and allows one to specify options, such as regulatory control and molecular structures.	4.0
F38	Colourful and/or 3D.	3.5
F39	Provide overview that is organized in multiple ways (e.g. organism, tissue, cell type, cell compartment).	3.5
F40	Provide links to literature, updated every year (e.g. PubMed).	3.0
F41	Ability to download or focus on a specific areas, for teaching purposes	2.0
F42	Customizable (i.e. starts simple and gets more complex).	2.0

Feature #	Feature	Votes
F43	Provide ability to have layers where each layer is simpler (e.g. "Just Enzymes").	2.0
F44	Provide ability to edit and add pathways (for personal use).	2.0
F45	Provide links to "all" basic information, such as diseases and all names of an enzyme.	1.0

Note that the reason for having fractions in the total number of votes is because one user gave one vote for both features F33 and F39, and one vote for both features F36 and F38. Thus, these votes were split in half between the features, and each feature was given 0.5 votes.

As Table 3-5 shows, the most important features are visualizing dynamic relationships, displaying an overview of biochemical pathways, and developing an easy to use, comprehensive, captivating, and interactive tool, with a good search engine. Having a hierarchical organization according to organ, tissue, etc, and the use of colors/3D received three and a half votes. The rest of the features received two votes, except having links to all basic information, which received one vote. For displaying an overview of pathways, users commented that they would like to be able to see the interactions between pathways, and they would like to be able to zoom-in to see more details. They also would like to see how they got there on a side bar.

Other features that users thought are important (from a teaching perspective) and were mentioned at some point in the session are cutting and pasting (for embedding information into other documents), printing a pathway figure along with the legend, and animation. Users thought that animation would allow users to better understand biochemical pathways. Examples of complex animations are animation of reactions (e.g. removing the hydroxyl group) and animation of traffic of molecules across membranes. An example of simpler animations is highlighting a pathway.

Table 3-6 summarizes the most difficult tasks users identified when using biochemical pathways that a computerized tool may be able to support.

Table 3-6: A list of most difficult tasks with biochemical pathways, found by 12 users in the brainstorming session

Difficulty	Description
D1	Knowing credibility (i.e. knowing where the information came from so as to be able to assess its reliability)
D2	Finding tissue-specific pathways.
D3	Finding pathways catalyzed by the same enzyme.
D4	Finding the relevance or importance of a reaction or pathway.
D5	Remembering "key features" (e.g. by using visual aids).
D6	Assessing the extent to which information is current or up-to-date, and knowing when was it updated.
D7	Seeing older versions and knowing the rationale for changes.
D8	Being able to download or copy and paste it (e.g. to place it in a PowerPoint presentation).
D9	Performing inter-species comparison.

From the description of the difficulties, one can see that users would like to be able to search for pathways that are specific to a certain tissue or organism, or that contain a specific enzyme. Finding information about a pathway, such as updates, source (credibility), and relevance, is also important. In addition, having visual aids, such as color, is important to help users remember key features of pathways. Lastly, for teaching purposes, professors would like to be able to download a pathway (perhaps save it as a jpeg or a vector format) in order to be able print it or incorporate it in some other program.

With respect to searching, users said they would like to be able to do single and multiple searching, as well as quick and refined searching. Users said they would like to search for pathways by species, tissue, cell type, and name, as well as by the enzymes, compounds, macromolecules (e.g. proteins), cofactors, regulators (activators and inhibitors), and reactions they contain. When asked about what users would like to have in the quick search, most agreed that prompting for species and pathway name is sufficient, while enzymes, tissues, and cell types would be needed only in a more detailed search facility. However,

others thought that enzymes are more important for quick searching, and that species are less important, because they use the former more for research.

When asked if users would like to be able to use wild card truncation in the search engine, they commented that they would like to, but that it was not very important.

With respect to classifications and representation of pathways, users commented that they like the metabolic/regulatory pathway classification, and with respect to metabolic pathways, the classification according to the macromolecules metabolized. Users also commented that they would like to see a tree view of pathways, where clicking on a pathway gives a menu of options, such as displaying a pathway graph. Users also commented that for enzyme names, the tool should use their common names in the search, while providing their formal name and E.C. number when clicking on (or hovering over) the enzyme to display additional information.

For representing pathways, users preferred being able to choose having more or less detail in the pathway graph, similar to MetaCyc. For instance, users can choose to only have the compounds' names or their names and chemical structures, depending on the level of detail chosen. Users commented that they would like to be able to compare pathways, but that the use of multiple windows adds complexity. Perhaps the use of tabbed panel could solve this problem. Users thought that ExPASy is very complex and that manipulating the overview window by zooming in and out causes too much cognitive thinking. They wanted to see everything in a simplified manner, perhaps displaying one pathway at a time, instead of having several pathways in the same window, which adds complexity. If we choose to show several pathways in the same window, then perhaps the number should be limited to three at a time. Users also commented that they would like to see a legend underneath the pathway figure, where they can scroll down to see it, perhaps having a choice of expanding the legend or hiding it. The reason for having the legend underneath the figure is to be able to have the option of printing it along with the figure.

Finally, users spent some time discussing 3D versus 2D for the visualization of biochemical pathways. Some users thought that 3D is not necessary because it is much more complicated, very difficult to handle, and requires intellectual effort. Other users commented that 3D would be an interesting topic to investigate and suggested using the third dimension to see interconnections between different layers of pathways, where each layer represents a class of pathways (i.e. carbohydrate metabolism in one layer, lipid metabolism in another layer, etc). Basically, the idea is to use depth to give a perception of 3D.

Chapter 4 BioPathVis, the Biochemical Pathway Visualization Tool, Designed Based on User Preferences

In this chapter, we will discuss the design and the main components of BioPathVis, the biochemical pathway visualization tool, designed based on user preferences described in Chapter 3, while taking into account the usability issues discussed in Chapter 2. We will start with the use cases, which translate into requirements. Then, we will discuss the system overview and architecture, as well as the major UI design decisions we faced. Subsequently, we will present the BioPathVis prototype. Finally, we will present a comparison of the feature set provided by BioPathVis to existing tools.

4.1 Use Cases

A use case is described as a snapshot of one aspect of a system [11]. The sum of all use cases is the external picture of the system [11]. Hence, a collection of use cases represent the requirements gathered from users.

Based on the user studies conducted in Chapter 3, the following are the actors and the user cases that are desired in a biochemical pathway visualization tool.

Actors:

1. Referencer (represents students and professors who are solely interested in visualizing biochemical pathways).
2. Researcher (represents researchers who are interested in both visualizing and editing biochemical pathways).

Desired Use Cases:

1. *Display overview of pathways.*
2. *Search for pathway(s) by name.*
3. *Search for pathway(s) by* organisms.
4. *Search for pathway(s) by class.*
5. *Search for pathway(s) in different organisms, tissues, organs, cells, and organelles.*
6. Search for pathway(s) containing reaction(s).
7. Search for pathway(s) containing compound(s).
8. *Search for reaction(s) by name.*
9. *Search for reaction(s) by* organisms.
10. *Search for reaction(s) by type.*
11. *Search for reaction(s) in different organisms, tissues, organs, cells, and organelles.*
12. Search for reaction(s) by pathway(s).
13. *Search for compound(s), other than enzymes, by name.*
14. *Search for enzyme(s) by name.*
15. *Search for compound(s) by* organisms.
16. *Search for compound(s) by type.*
17. *Search for compound(s) in different organisms, tissues, organs, cells, and organelles.*
18. Search for compound(s) by pathway.
19. *Quick search for compound(s) in pathway figure.*
20. *Present info by clicking on compound in pathway figure.*
21. *Present info by clicking on enzyme in pathway figure.*
22. *Present info by clicking on reaction in pathway figure.*
23. *Zoom in and out of pathway figure.*
24. *Show/hide details in pathway figure.*
25. *Display adjacent pathway links.*
26. *Display legend.*
27. *Present consistent color coding for compounds.*
28. Provide accessible help menu.
29. *Display tool tips.*
30. Display interactive tutorial about tool.

31. View glossary.
32. Display compound's 3D structure.
33. Include iso-enzymes.
34. View links to literature.
35. View updated date of pathway.
36. View older versions of pathway.
37. Save pathway figure as jpeg.
38. Print pathway figure along with legend.
39. Add/edit pathway(s) dynamically.
40. Add/edit reaction(s) dynamically.
41. Add/edit compound(s) dynamically.
42. Provide capability to roll back to previous state.
43. Provide clear description of current state (e.g. enable buttons that can be clicked).
44. Provide user's own database.
45. Visualize dynamic relationships.
46. Provide customizable features.
47. Provide animations.

Note that a compound represents any compound present in the pathway, including enzymes, macromolecules, cofactors, and regulators. The reason for having use cases 13 and 14 is that some existing tools only provide the capability to search for enzymes, and not other kinds of compounds.

Because of time limitations, only the use cases in italics, which are considered the most important (according to Figure 3-9, Table 3-3, Table 3-5, Table 3-6) and easiest to implement, were implemented.

4.2 Configuration and Tools

The BioPathVis tool was written in Java. The data used in the BioPathVis application was populated in the Oracle database.

The following tools were used to develop and run the BioPathVis application.

1. Java SDK 1.4 (Java development toolkit, containing essential packages to develop the application).
2. Oracle 9i.
3. JBuilder 7.0 (A tool to develop the Java files).

The following data is required before BioPathVis can connect to the Oracle database:

1) The Database Management System (DBMS) location (hostname and port).
2) The database name.
3) The user name to log in as.
4) The password for the user.
5) The driver required for the connection to the Oracle database.

4.3 High Level Architecture and Design

4.3.1 System Overview

Because the purpose of our study is to design the UI of a biochemical pathway visualization tool, we will focus only on the high level design of the tool, in order for others to be able to duplicate this work more easily.

BioPathVis is comprised of three main layers: The DbInterface layer, the Graph layer, and the GUI Layer. The DbInterface layer is responsible for connecting to the Oracle database through JDBC and for handling all database operations. JDBC is a platform-independent, database-independent set of classes written in Java that allow Java programs to connect to

databases. It is the standard mechanism provided by Sun for database connectivity in the Java language. JDBC cannot connect to a DBMS by itself. Before JDBC can connect to a DBMS, it requires a driver designed for that DBMS. The DbInterface layer can be thought of as an abstraction between the rest of the BioPathVis application and the JDBC bindings that are used by the Java language to connect to the DBMS. This abstraction allows the database connection mechanism to change in the future. The DbInterface layer provides an interface to the GUI layer through the DbInterface class.

It is worth mentioning that the choice of the Oracle database is purely based on the fact that data was already populated in Oracle by IfCOS and made available for us to use for the purpose of this research. Had this data not been available, we would have used a less expensive form, such as XML. In fact, the tool supported the use of XML files to read data. However, data was not yet populated into these XML files in the format specified to be used with the tool.

As was mentioned in section 2.2.3, object models will be used to represent biochemical pathways. Hence, data read from the database is populated into objects with attributes. These objects are Species, Organ, Tissue, Cell, Organelle, Pathway, Reaction, Compound, and Enzyme objects, which are also part of the DbInterface layer.

The Graph layer is responsible for handling all operations related to drawing and laying out biochemical pathway graphs. The main library used in the Graph layer is JGraph, which is an open source interactive Java graph visualization library containing graphical and algorithmic functionality [15]. There are other graph libraries that contain algorithms that are more specific for biochemical pathways, such as yWays [41]. However, these libraries require the purchase of a license. The main reason for choosing JGraph is because it is open-source and serves our purposes for creating acceptable graph layouts.

The GUI layer provides a graphical user interface for the user to query the database, through the DbInterface layer, and to visualize pathway graphs, through the Graph layer.

Figure 4-1 below is a package diagram of the main packages in the BioPathVis application, which better illustrates the dependencies among the different components. The packages that are developed for the purpose of this study are the gui, graph, and dbinterface packages. Other packages, such as javax.swing, org.jgraph, oracle.jdbc, and oracle.sql are third party packages that help in developing the application. Many other packages, such as java.awt and java.io, were used. However, these are not included in the diagram for clarity.

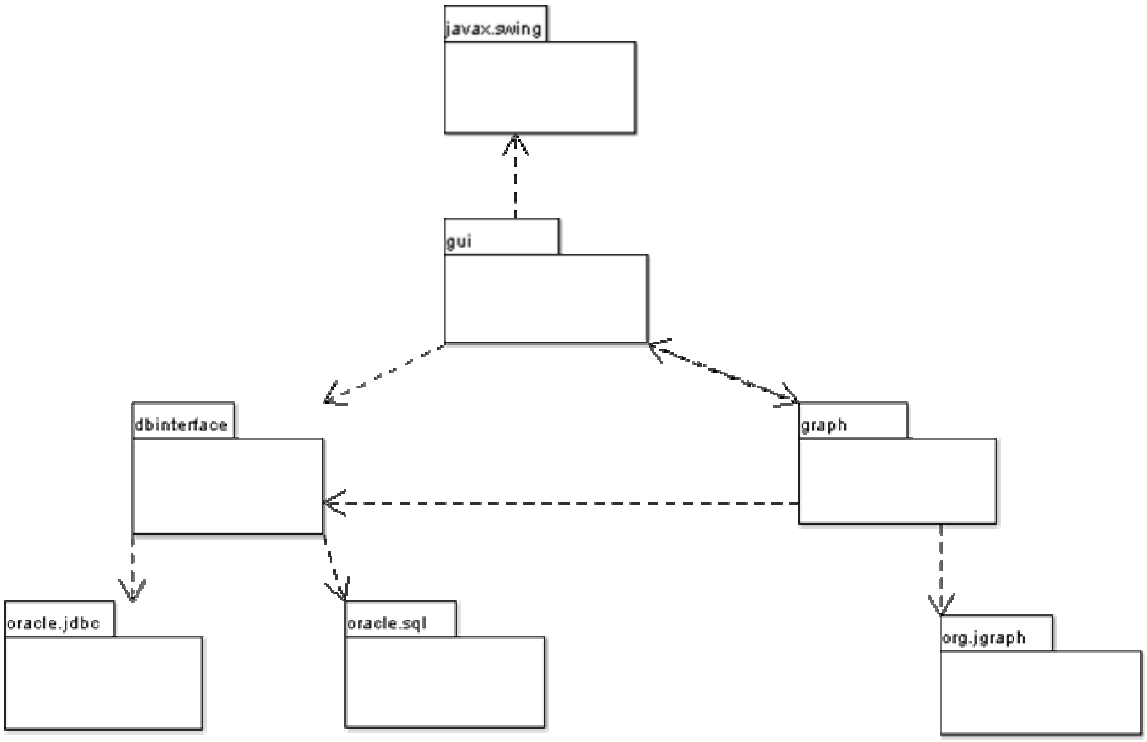


Figure 4-1: BioPathVis package diagram

As we can see in Figure 4-1, the dbinterface package depends only on packages that are necessary to connect to the Oracle database.

The graph package depends on the gui, dbinterface, and org.jgraph packages. As was mentioned earlier, the graph package uses the JGraph library encapsulated in the org.jgraph package, to draw pathway graphs. It uses the dbinterface package when the user attempts to retrieve information about the graph components (nodes and arcs) by clicking on them.

Finally, the graph package uses the gui package if such user actions require popping up dialog boxes or changing the state of menu items, which are part of the gui package.

The gui package depends on the dbinterface, graph, and javax.swing packages. The gui package uses javax.swing to draw all the necessary components of the user interface. It also passes user actions into the dbinterface layer, which subsequently queries the database. When the results returned from the database are graphical in nature, the gui package requests the graph package to draw the graph.

4.3.2 Layout Algorithm

As was mentioned earlier, in order to display a dynamic biochemical pathway graph, the pathway data is read from the database and populated in the pathway graph. The object model, described in section 2.2.3, where the nodes and edges are objects, is employed in BioPathVis. There are 3 types of nodes in pathway graphs drawn using BioPathVis: Metabolite nodes represent the substrates, products, and enzymes, adjacent pathway nodes represent pathways that are adjacent to the current pathway, and reaction nodes represent the reactions.

As was mentioned in section 4.3.1, the JGraph library is used to manipulate pathway graphs. The nodes and edges are added to the pathway graph dynamically as the pathway data is read from the database and the Spring Embedded Layout algorithm is used to layout the final pathway graph. The reason for choosing the Spring Layout Algorithm is that it produced acceptable layouts for biochemical pathways as compared with other algorithms provided by the JGraph library. Although the Spring Embedded Layout algorithm is a good starting point for laying out the pathway graphs, it nevertheless produces layouts that are hard to read because of edge crossings and nodes placed on top of one another. Hence, after applying the layout algorithm, the pathway graphs are laid out manually and saved as .pathway files. When displaying a pathway graph, if the .pathway file corresponding to that pathway exists, then the pathway information is read from the file. Otherwise, the pathway information is

read from the database and the default Spring Embedded Layout algorithm is applied. In the future, we will create a more customized drawing algorithm to draw biochemical pathways.

4.4 UI Design Decisions

While designing the UI of BioPathVis, we were faced with several choices for designing certain UI features. In this section, we will highlight these main UI design decisions and the rationale behind these decisions.

The first is the use of tree view to display pathway graphs. Biochemical pathways are organized into metabolic and regulatory pathways, and within each category, the actual pathways are organized into classes, such as amino acid metabolism, carbohydrate metabolism, etc (See section 1.2). Because biochemical pathway information is hierarchical in nature, it was only natural to decide to use a tree view to organize such a hierarchy. When users were asked about pathway organization in the brainstorming session, they commented that they liked the tree view organization (See section 3.3.2).

The second UI design decision is the use of a tabbed panel to display static overview images as well as dynamic individual pathway graphs. Initially, we thought of using multiple windows to represent biochemical pathways, where one window displays an overview, and several other windows display the actual pathways. The use of multiple windows adds complexity to the interface, as shown in Section 2.3.2.4. Hence, we considered an alternative solution, which is to use a tabbed panel to display static overview images as well as dynamic pathway graphs. The use of a tabbed panel has its disadvantages when the number of tabs becomes huge. However, we assume that users would simply close tabs that they don't need and that users would not need to display such a huge number of tabs simultaneously. Because of time limitation, we were not able to conduct more tests to validate our assumption. However, if more tests are conducted and these tests show that our assumption is invalid, we can simply use navigation buttons along with the tabbed panel. These buttons

would hide additional tabs if they don't fit into one line and would allow the user to fast forward from one tab to the next.

The third UI design decision is the use of static KEGG images to display overview images of pathways. Since static KEGG overview images were readily available as GIF files, and because of their nice colors and ease of incorporation in BioPathVis, we decided to use these images as overview images as opposed to accessing the database to draw these overview graphs dynamically. With the use of KEGG static images, the user is presented with an overview image of all pathway classes. When the user double clicks on a pathway class, it shows another overview image of all pathways that belong to that class. Double clicking on a pathway shows the actual pathway as a new tab (this pathway is drawn dynamically as it is read from the database using JGraph).

The alternative (and ultimate) solution to draw overview images is to draw them dynamically the same way individual pathway graphs are drawn. Using this solution, pathway class nodes are read from the database and drawn in the overview graph. One pathway class node is connected to another if it contains at least one pathway that is connected to at least one other pathway in the other class node. If the user double clicks on a class node, it will expand to show all pathways it contains (i.e. using Elision techniques – see Section 2.3.2.3). Although this solution is the most desirable, it is more time consuming than using KEGG static images, as more APIs have to be written to access the data source and to get the adjacent pathway links. These APIs must return all pathway links so that these links can be drawn using JGraph. This solution should be implemented later as an enhancement to BioPathVis and a replacement of the existing solution.

The fourth UI design decision is related to printing pathway graphs. One of the features users mentioned they would like to have in a biochemical pathway visualization tool is to be able to print a biochemical pathway. One way to accomplish this task is to allow the user to convert the graph into a vector format and print it. Another way is to allow the user to save the graph as a JPEG or GIF file and then print that file. Because the latter alternative is easier

to implement, given the fact that saving a graph is supported in the JGraph library, we decided to use this alternative.

The fifth UI design decision is related to which items should be incorporated in the quick and detailed searching. Users commented that they would like to be able to search for pathways in different species, organs, tissues, cells, and organelles. We thought that it would be best to incorporate this capability in the quick search, as items just above the tree view of pathways. We believe that this decision would make the tool more extensible, for instance, for future incorporation of simulation capabilities. For simulation, users would need to be able to select different models, which contain organs, tissues, cells and organelles. Model selection would easily be added as an item along with the other pathway hierarchy items. Also, because we use a tree view to display pathways in the quick search, we thought that it is not important to provide searching for pathway name by typing a partial pathway name in the quick search. We thought that this feature should go in the detailed search, by offering a text field to type in partial pathway name; this takes user to the closest matching pathway name in a list of pathway names located just underneath the text field.

The sixth UI design decision is to use a tabbed panel to represent the legend, for extensibility. For instance, we would have one tab to describe the legend for visualization, another later on for editing, and yet another for simulation. Although this approach does not allow the user to print the legend along with the pathway, it allows for extensibility of the tool by adding more information in an organized manner. Later on, we can add a feature that allows the user to print the information in the legend. We believe that for visualization purposes, the user would need the legend only when starting to use the tool. Later on, the user would become familiar with the colors and would know what the nodes mean by just considering their content (i.e. name) without looking at the color. For instance, for biochemists, they would know that “Glucose” is a carbohydrate regardless of its color.

4.5 BioPathVis Prototype

Figure 4-2 below is a screen shot of the BioPathVis application upon startup.

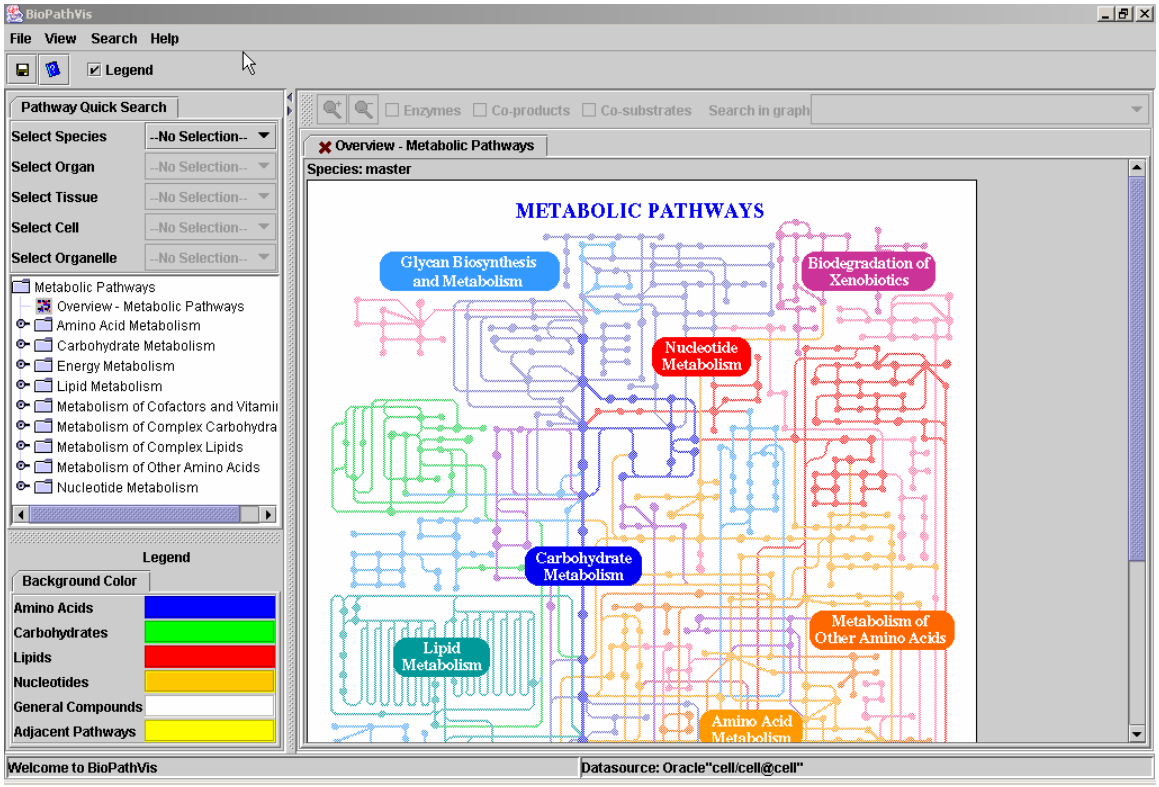


Figure 4-2: BioPathVis upon startup

As shown in Figure 4-2, the BioPathVis application consists of a menu, a toolbar, a status bar, a pathway quick search panel located on the top left corner, a legend located on the bottom left corner, and a pathway display area in the center, with a navigation bar on top.

The pathway quick search panel allows for a quick hierarchical search for pathways in different species, organs, tissues, cells, and organelles. By default, the general (or "master", a term used in KEGG [20]) pathway names are displayed in a tree view below the combo boxes. Selecting a species displays pathway names belonging to that species and enables the organ, tissue, cell, and organelle combo boxes. Similarly, selecting an organ, tissue, cell, and/or organelle displays pathway names belonging to these selections.

The tree view presents pathways by their classes as the parent nodes (i.e. “Amino Acid Metabolism”, “Carbohydrate Metabolism”, etc) and their names as leaf nodes (i.e. “Glycolysis / Gluconeogenesis”). The first leaf node of each pathway class parent node is an overview leaf. Double clicking on the overview leaf displays a static overview image of pathways that belong to that class, as a tab in the pathway display area. The subsequent leaf nodes of each pathway class node are the actual pathways. Double clicking on any of these pathway leaf nodes displays a dynamic pathway graph that corresponds to the leaf node name, as a tab in the pathway display area. Figure 4-3 below shows a pathway graph for the “Nucleotide sugars metabolism” pathway in the “E.coliK-12” species. In Figure 4-3, the pathway display area is expanded horizontally to fit the entire screen. To expand the pathway display area, the user needs to place the mouse on the divider that divides the pathway quick search area from the pathway display area and hold the mouse left button while dragging the mouse to the left.

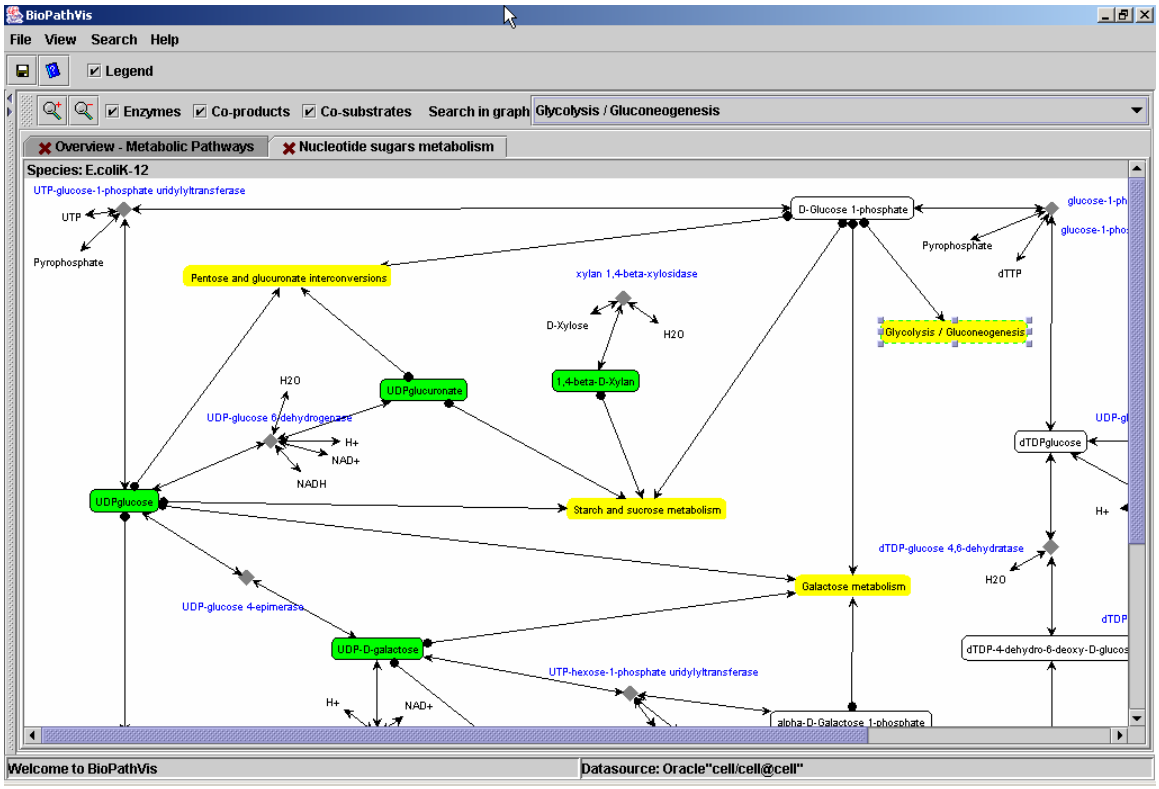


Figure 4-3: Representation of “Nucleotide sugars metabolism” pathway in “E.coliK-12” species, using BioPathVis. Pathway display area is expanded horizontally to fit entire screen

The pathway figure above consists of nodes and edges. Borderless round-rectangular nodes in yellow represent adjacent pathways. The rest of round-rectangular nodes, which have borders, represent the various substrates and products in the pathway. The color coding is defined in the legend in Figure 4-2. Non-filled, borderless nodes in blue text represent enzymes, whereas non-filled, borderless nodes in black represent co-substrates and co-products. The reason for using such a representation is that it mimics the representation and colors used in Michal's "Biochemical Pathways" [28], which is familiar to biochemists. Reaction nodes are represented as diamond nodes in gray, to distinguish them from other nodes in the pathway diagram.

Double clicking on an adjacent pathway node displays the pathway graph of the adjacent pathway, as a tab in the pathway display area. Double clicking on any other compound or reaction node displays information about the node, as shown in Figure 4-4 below.

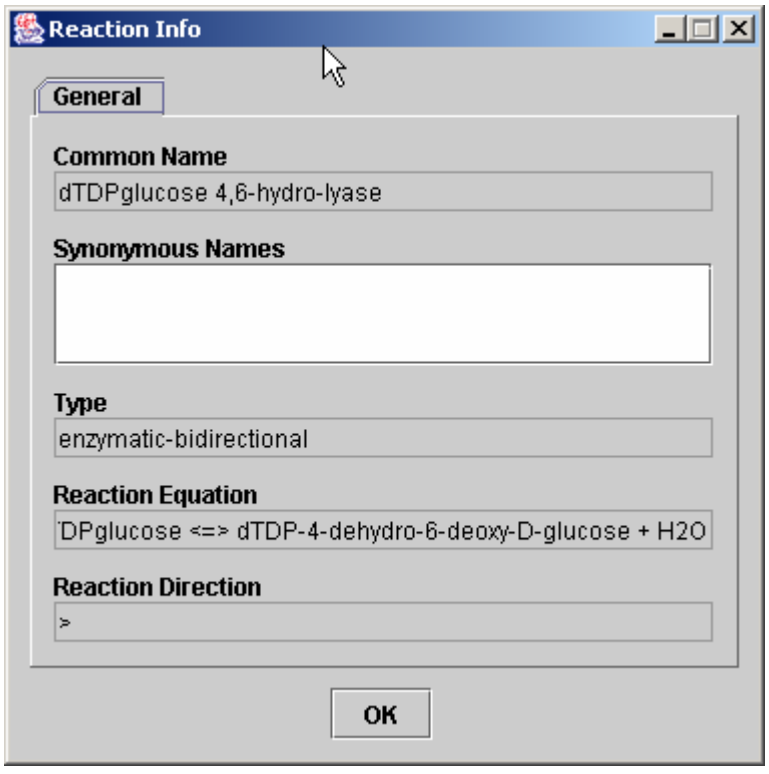


Figure 4-4: "Reaction Info" dialog box, which results from clicking on "dTDPglucose <=> dTDP-4-dehydro-6-deoxy-D-glucose + H2O" reaction node in "Nucleotide sugars metabolism" pathway in "E.coliK-12" species

As shown in Figure 4-3, there is a pathway navigation status bar on top of the pathway display area. This pathway navigation bar allows the user to zoom in and out of the pathway figure, to hide enzymes, co-products, and/or co-substrates, and to search for information in the pathway graph. For instance, if the user scrolls down the "Search in graph" combo box to select "Glycolysis / Gluconeogenesis", the pathway graph scrolls to the "Glycolysis / Gluconeogenesis" node (if that node is not visible) and selects it in green. Figure 4-5 below shows the same pathway graph zoomed out to fit the entire screen, and with enzymes and co-substrates hidden.

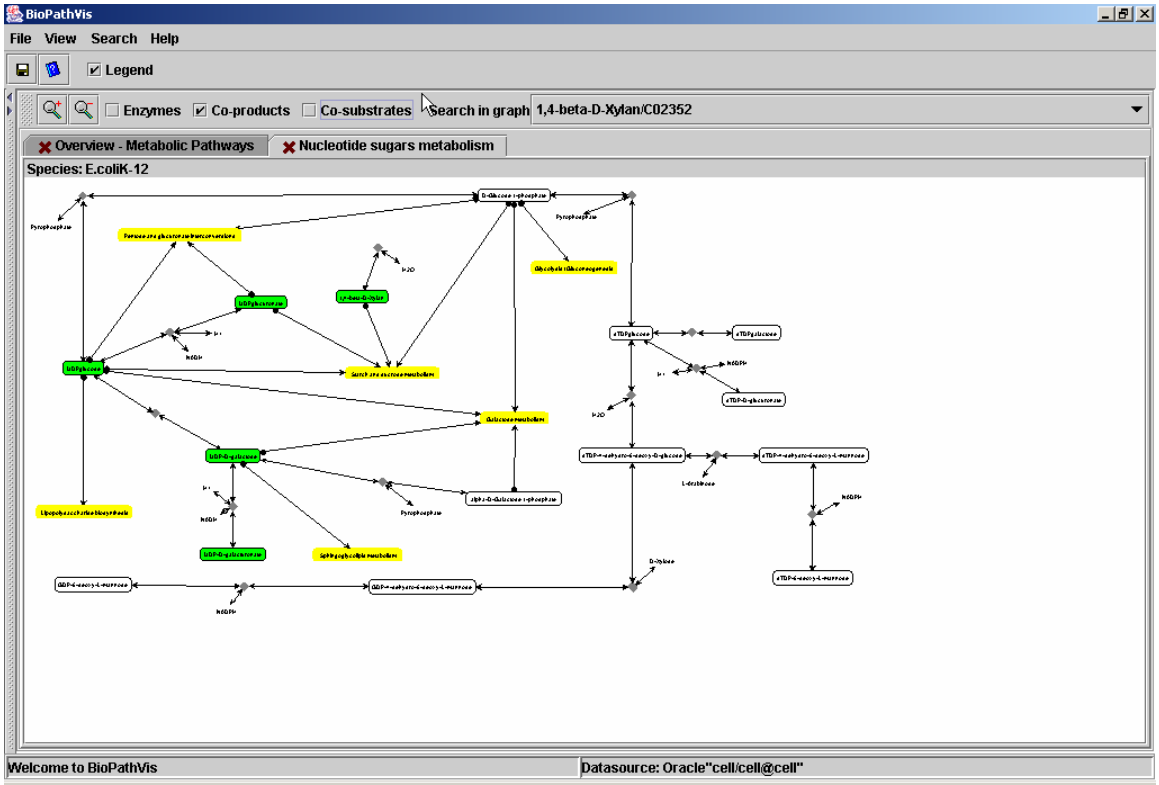


Figure 4-5: Representation of "Nucleotide sugars metabolism" pathway in "E.coliK-12" species, zoomed out, and with enzymes and co-substrates hidden. Pathway display area is expanded horizontally to fit entire screen

We will now describe the menu of BioPathVis. The menu consists of the "File", "View", "Search", and "Help" menu items. The "File" menu consists of the "Save As" item, which allows the user to save a pathway image or overview as a jpeg file (See Figure 4-6 below), and the "Exit" item which allows the user to quit the application.

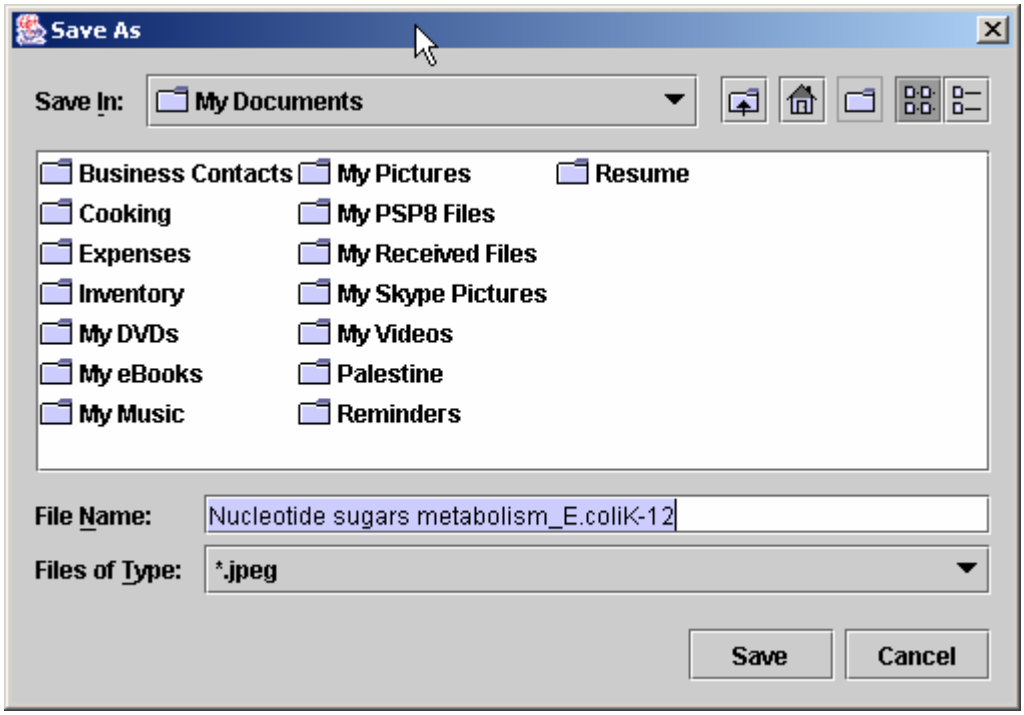


Figure 4-6: "Save As" dialog box. Default pathway name is automatically entered in the "File Name" field

The "View" menu is shown in Figure 4-7 below. The first two items allow the user to hide or show the tool bar and status bar. They are currently checked because the tool bar and status bar are shown in the tool. The "Pathway Overview" item allows the user to display a static image of metabolic pathways. It is currently disabled because a pathway overview image is already shown (See Figure 4-2). The "Legend" check box allows the user to show or hide legend. The "Zoom in" and "Zoom out" items allows the user to zoom in and out of a pathway graph. These are disabled if the selected image is a static overview image.

Figure 4-7: "View" menu

The "Search" menu allows the user to perform detailed searching. It consists of the "Search for Pathway(s)…", "Search for Reaction(s)…" and "Search for Compound(s)…" items. Clicking on the "Search for Pathway(s)…" item displays a dialog box that allows users to perform advanced searching for pathways, as shown in Figure 4-8 below. At the current time, this dialog box consists only of a "General" tab. Later, more advanced searching tabs will be added, specifically ones that allow for searching for pathways containing certain reactions and compounds. The "General" tab allows the user to search for pathways in a similar manner as the quick search. Additionally, it allows the user to select which pathway classes the user wishes to find. It also allows the user to type in a part of the pathway name, which takes the user to the closest matching pathway name. The user can select multiple pathways by holding down the 'Ctrl' or 'Shift' keys while making the selections.

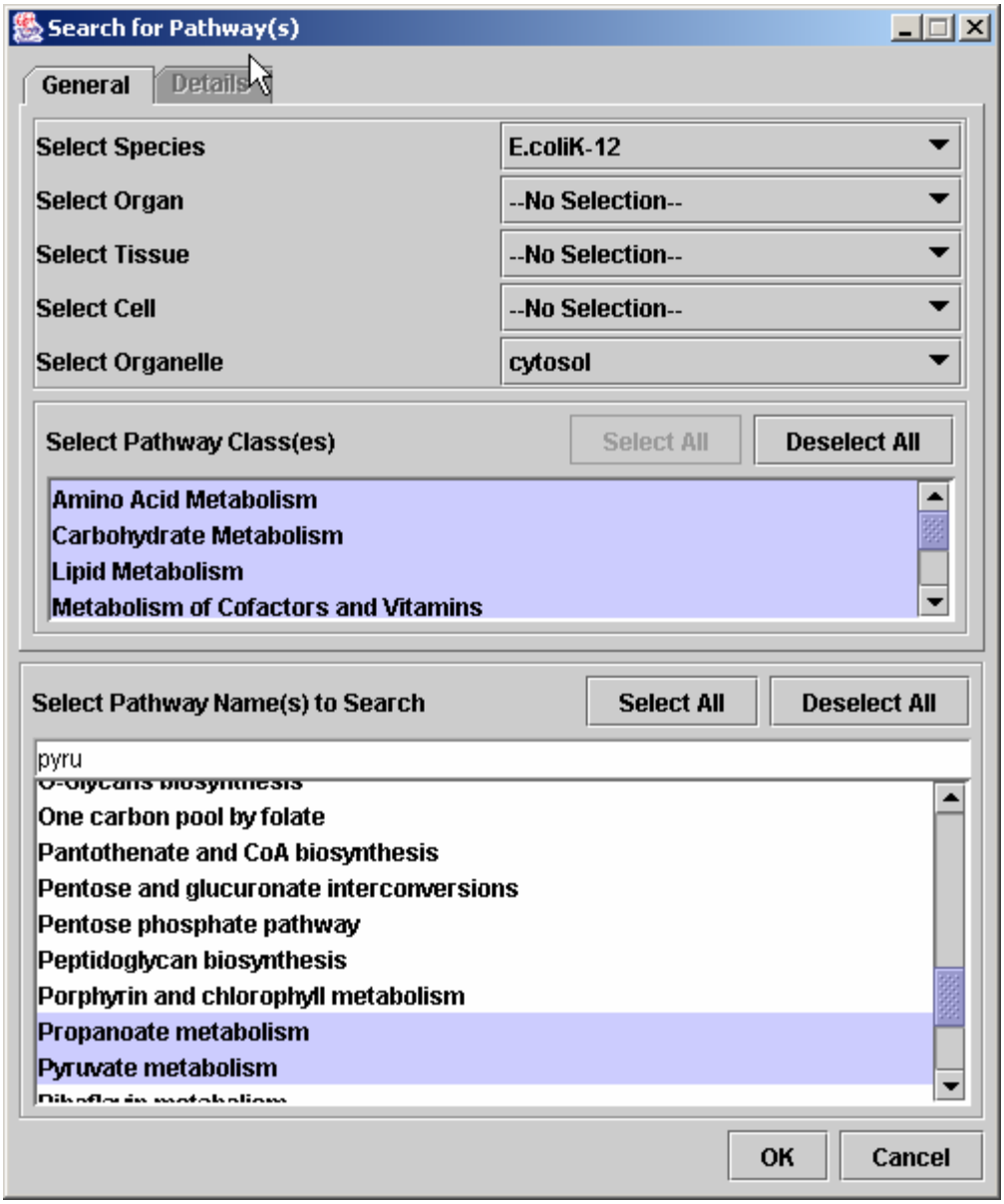


Figure 4-8: Advanced searching for pathways. Here, in E.coliK-12 species, the user wishes to find simultaneously "Pyruvate metabolism", "Propanoate metabolism" and "Nucleotide sugars metabolism" pathways

Clicking on the "Ok" button displays the search results as a tab in the search panel, besides the quick search tab, as shown in Figure 4-9 below. The pathway search results tab behaves in a similar manner as the quick search tab, except for the fact that the user cannot modify the species, organ, tissue, cell, and organelle because the user has already chosen these using the advanced search.

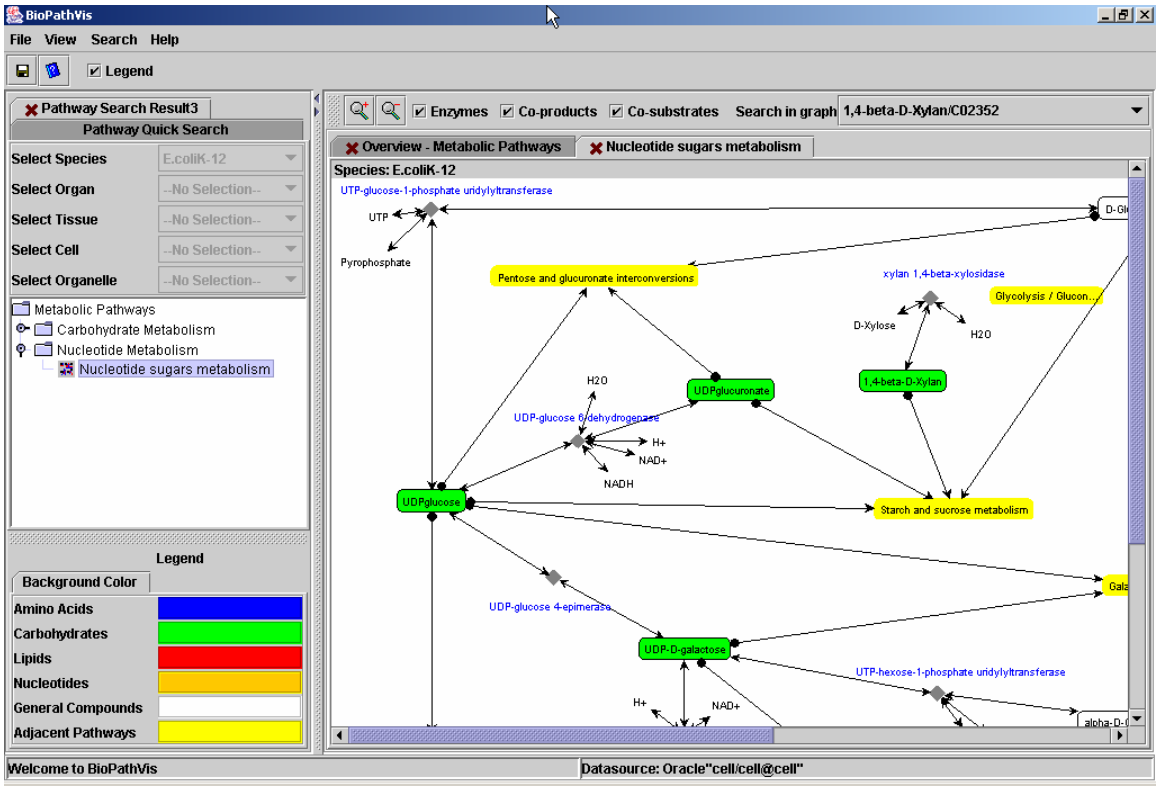


Figure 4-9: Advanced pathway search results, after clicking on "Ok" button

Searching for compounds and reactions is done in a similar manner. Searching for reactions allows the user to select reaction classes, such as "enzymatic-bidirectional". Searching for compounds allows the user to select compound types, such as "lipid" and "amino acids".

The "Help" menu consists of 1 item: "About BioPathVis…" which displays information about the BioPathVis application. In the future, this menu will consist of various help topics to help users with the BioPathVis application.

From the description provided above, the BioPathVis application clearly covers all use cases presented in italics in section 4.1.

4.6 Feature Set Provided by BioPathVis versus Existing Tools

Table 4-1 below compares the set of features provided by BioPathVis versus other tools. The "Related Feature / Difficulty" column is obtained from Figure 3-9, Table 3-3, Table 3-5, and Table 3-6. An average importance value, on a scale from 1 to 5, where 1 is "Very unimportant" and 5 is "Very important", is assigned to each feature based on Figure 3-9 and Table 3-3. For each tool, an existence value of 0 is given to the tool if the feature does not exist and a value of 1 is given to the tool if the feature exists. The weight is calculated by multiplying the average importance value by the existence value. The total weight of all features in a tool is calculated by adding the weights of all features. Please note that "BC" stands for BioCarta, "E" for ExPASy, "K" for KEGG, "M" for MetaCyc, "W" for WIT and "BP" for BioPathVis.

Table 4-1: Feature set provided by BioPathVis versus existing tools

FS #	Related Feature / Difficulty	Description	Avg. Imp.	Weight					
				BC	E	K	M	W	BP
FS1	*F3, F33*	*Display overview of pathways.*	*4.3*	*0.0*	*4.3*	*4.3*	*4.3*	*0.0*	*4.3*
FS2	*F1, F34*	*Search for pathway(s) by name.*	*4.6*	*4.6*	*0.0*	*4.6*	*4.6*	*0.0*	*4.6*
FS3	*F7, F34, D9*	*Search for pathway(s) by organisms.*	*3.9*	*3.9*	*0.0*	*3.9*	*3.9*	*0.0*	*3.9*
FS4	*F1, F34*	*Search for pathway(s) by class.*	*4.6*	*0.0*	*0.0*	*4.6*	*4.6*	*0.0*	*4.6*
FS5	F22, F34, F39, D2, D9	Search for pathway(s) in different tissues, organs, cells, and organelles.	5.0	0.0	0.0	0.0	0.0	0.0	5.0
FS6	*F6, F34*	*Search for pathway(s) containing a reaction.*	*4.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
FS7	*F4, F34, D3*	*Search for pathway(s) containing a compound.*	*4.3*	*0.0*	*4.3*	*0.0*	*4.3*	*4.3*	*0.0*
FS8	*F2, F34*	*Search for reaction(s) by name.*	*4.3*	*0.0*	*0.0*	*4.3*	*4.3*	*0.0*	*4.3*

FS #	Related Feature / Difficulty	Description	Avg. Imp.	Weight					
				BC	E	K	M	W	BP
FS9	*F7, F34, D9*	*Search for reaction(s) by organisms.*	*3.9*	*0.0*	*0.0*	*0.0*	*3.9*	*0.0*	*3.9*
FS10	*F2, F34*	*Search for reaction(s) by type.*	*4.3*	*0.0*	*0.0*	*0.0*	*4.3*	*0.0*	*4.3*
FS11	F22, F34, F39, D2, D9	Search for reaction(s) in different tissues, organs, cells, and organelles.	5.0	0.0	0.0	0.0	0.0	0.0	5.0
FS12	*F2, F34*	*Search for reaction(s) by pathway.*	*4.3*	*0.0*	*0.0*	*4.3*	*4.3*	*0.0*	*0.0*
FS13	*F5, F34*	*Search for compound(s) (non enzymes) by name.*	*4.1*	*0.0*	*0.0*	*4.1*	*4.1*	*0.0*	*4.1*
FS14	F16, F34	Search for enzyme(s) by name.	5.0	5.0	5.0	5.0	5.0	5.0	5.0
FS15	*F7, F34, D9*	*Search for compound(s) by organisms.*	*3.9*	*0.0*	*0.0*	*0.0*	*3.9*	*0.0*	*3.9*
FS16	*F5, F34*	*Search for compound(s) by type.*	*4.1*	*0.0*	*0.0*	*0.0*	*4.1*	*0.0*	*4.1*
FS17	F22, F34, F39, D2, D9	Search for compound(s) in different tissues, organs, cells, and organelles.	5.0	0.0	0.0	0.0	0.0	0.0	5.0
FS18	*F5, F34*	*Search for compound(s) by pathway.*	*4.1*	*0.0*	*0.0*	*4.1*	*0.0*	*0.0*	*0.0*
FS19	*F5*	*Quick search for compound(s) in pathway figure.*	*4.1*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*4.1*
FS20	F19	Present info by clicking on compound in pathway figure.	5.0	0.0	0.0	5.0	5.0	0.0	5.0
FS21	F21	Present info by clicking on enzyme in pathway figure.	5.0	5.0	0.0	5.0	5.0	5.0	5.0
FS22	F19	Present info by clicking on reaction in pathway figure.	5.0	0.0	0.0	0.0	5.0	0.0	5.0
FS23	F23	Zoom in and out of pathway figure.	5.0	0.0	0.0	0.0	0.0	0.0	5.0
FS24	F43	Show/hide details in pathway figure.	-	-	-	-	-	-	-

FS #	Related Feature / Difficulty	Description	Avg. Imp.	Weight					
				BC	E	K	M	W	BP
FS25	F17	Display adjacent pathway links.	4.5	0.0	4.5	4.5	4.5	0.0	4.5
FS26	F13	Display legend.	5.0	5.0	0.0	0.0	5.0	0.0	5.0
FS27	F15, F38	Present consistent color coding for compounds with a legend.	3.5	0.0	3.5	0.0	0.0	0.0	3.5
FS28	F11	Provide accessible help menu.	5.0	0.0	0.0	0.0	0.0	0.0	0.0
FS29	F11	Display tool tips.	5.0	0.0	0.0	0.0	0.0	0.0	5.0
FS30	F12	Display interactive tutorial about tool.	4.0	0.0	0.0	0.0	0.0	0.0	0.0
FS31	F14	View glossary.	3.0	0.0	0.0	0.0	0.0	0.0	0.0
FS32	F18, F28, F37	Display compound's 3D structure.	5.0	5.0	5.0	0.0	5.0	0.0	0.0
FS33	F30	Include iso-enzymes.	3.0	0.0	0.0	0.0	0.0	0.0	0.0
FS34	F27, F40, F45, D1, D4	View links to literature.	5.0	5.0	0.0	5.0	5.0	0.0	0.0
FS35	D6	View updated date of pathway.	-	-	-	-	-	-	-
FS36	D7	View older versions of pathway.	-	-	-	-	-	-	-
FS37	F20, F41, D8	Save pathway figure as jpeg.	5.0	0.0	0.0	0.0	0.0	0.0	5.0
FS38	F20	Print pathway figure.	5.0	0.0	0.0	0.0	5.0	0.0	0.0
FS39	*F8, F44*	*Add/edit pathway(s) dynamically using tool.*	*3.7*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
FS40	*F9, F44*	*Add/edit reaction(s) dynamically using tool.*	*3.7*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
FS41	*F10, F44*	*Add/edit compound(s) dynamically using tool.*	*3.6*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*	*0.0*
FS42	F24	Provide capability to roll back to previous state.	4.0	4.0	4.0	4.0	4.0	4.0	4.0
FS43	F25, D5	Provide clear description of current state (e.g. enable buttons that can be clicked).	4.0	0.0	0.0	0.0	0.0	0.0	4.0
FS44	F29	Provide users own database.	4.0	0.0	0.0	0.0	4.0	0.0	4.0

FS #	Related Feature / Difficulty	Description	Avg. Imp.	Weight					
				BC	E	K	M	W	BP
FS45	F31	Visualize dynamic relationships.	-	-	-	-	-	-	-
FS46	F42	Provide customizable features.	-	-	-	-	-	-	-
F47	-	Provide Animation.	-	-	-	-	-	-	-
FS48	F26, F32	Self-explanatory and easy to use.	-	-	-	-	-	-	-
FS49	F35	Comprehensive.	-	-	-	-	-	-	-
FS50	F36	Captivate the audience.	-	-	-	-	-	-	-
FS51	-	Modular and extensible, so that additional applications could be built on top of it.	-	-	-	-	-	-	-
Feature Set Count			**41.0**	**8.0**	**7.0**	**14.0**	**24.0**	**4.0**	**28.0**
Feature Set Total Weight			**178.8**	**37.5**	**30.6**	**62.7**	**103.1**	**18.3**	**121.1**
Percentage of Feature Set Total Weight			**100%**	**21.0%**	**17.1%**	**35.1%**	**57.7%**	**10.2%**	**67.7%**

The features provided in Table 4-1 represent features that are collected from user studies and the ones that are considered important and relevant to biochemical pathway visualization tools. The average importance values (and hence weight) of most of the features in Table 4-1 are determined directly from Figure 3-9 and Table 3-3. The exceptions are features FS4, FS9 – FS12, FS15 – FS19, FS29, and FS37, where the average importance values are deduced indirectly from the other features in Figure 3-9 and Table 3-3. Ideally, we would go back and validate our assumptions and deductions with users in another questionnaire, where we would ask the users to assign importance values for features that we do not have values for. However, given the time limitations and lack of availability of users, we were not able to do this.

Table 4-1 shows that the total feature set count determined from the user studies is 451 (since we exclude features that have no average importance values, as will be discussed shortly) and the total weight of all features is 178.8. When comparing the various biochemical pathway visualization tools, one can clearly see that BioPathVis has the highest feature set count (28), followed by MetaCyc (24), followed by KEGG (14), followed by BioCarta (8), followed ExPASy (7), and followed by WIT (4). In addition, BioPathVis has the highest total

feature set weight (121.1 or 67.7%), followed by MetaCyc (103.1 or 57.7%), followed by KEGG (62.7 or 35.1%), followed by BioCarta (37.5 or 21.0%), followed by ExPASy (30.6 or 17.1%), and followed by WIT (18.3 or 10.2%).

Note that the average importance values for features shown in red italics (FS1 – FS4, FS6 – FS10, FS12 – FS13, FS15 – FS16, FS18 – FS19) are determined from taking the opinion of 12 users on these features (See Figure 3-9), and hence the average importance values have better effect than the remaining features, which are determined from one or two users (See Table 3-3). Also, no average importance values were assigned to features FS24, FS35, FS36, or FS45 to FS51. The reason is that some of these features are obtained from Table 3-5 and Table 3-6, which do not provide any average importance values. In addition, features FS48 to FS51 are qualitative in nature and it's difficult to measure whether the tool provides them or not without additional experimentation.

If we only include features taken from Figure 3-9, we obtain 18 features of total weight of 73.8, MetaCyc would then have the highest feature set count (12), followed by BioPathVis (11), followed by KEGG (8), followed by BioCarta and ExPASy (2), and followed by WIT (1). In addition, MetaCyc would also then have the highest total feature set (50.6 or 68.6%), followed by BioPathVis (46.1 or 62.5%), followed by KEGG (34.2 or 46.3%), followed by ExPASy (8.6 or 11.7%), followed by BioCarta (8.5 or 11.5%), and followed by WIT (4.3 or 5.8%).

Chapter 5 Videotaped Evaluation of BioPathVis

5.1 Features Tested

All features of BioPathVis, described in sections 4.5 and 4.6 were evaluated. Although only features related to searching can be compared to other tools, the additional features that BioPathVis provides were tested to get as much malfunctions and feedback from users as possible.

5.2 Users

We attempted to get the same users that evaluated the existing biochemical pathway visualization tools. However, only 2 were available and hence we needed to look for other users.

Similarly to the videotaped evaluation with existing tools, we performed a pilot study with one user, who has a background in engineering and two years of experience with biochemical pathways. The purpose of the pilot study was to test the camcorder and the clarity of the questions. No modifications to the questions were necessary as a result of the pilot study, and hence the results were used for the videotaped evaluation.

Four additional users, with no previous knowledge or experience with BioPathVis, participated in this activity. These users were representatives of the target end-users since they have, at a minimum, a basic knowledge of biochemistry and computers. The first has a B.A. in biochemistry, the second has a B.A. in Bio-pharmacy, the fourth is a biochemistry lab technician, and the fifth is a chemist with a one year biochemistry background. Although

there was a lack of volunteers for testing BioPathVis, and despite the time constraints, the users chosen for the evaluation of BioPathVis represent a wide variety of users, with different biochemistry and computer backgrounds. Hence, the results are expected to vary based on the user's background.

5.3 Procedure

The videotaped evaluation was performed using the instructor's desktop (a DELL Dimension 8400 Intel(R) Pentium(R) 4 CPU 3.00GHz). Prior to the evaluation period, the users were informed of the purpose of the study and were asked to read and sign the informed consent forms (See Appendix B.1). The users were then given a quick PowerPoint tutorial on BioPathVis to ensure consistency with the videotaped evaluation of existing tools (See Appendix B.2). The tutorial consisted of 2 screen shots of BioPathVis describing it in general, and with no reference on how to use it.

The evaluation session took place at the evaluator's residence. Each session lasted from half an hour to an hour. A total of 6 tasks were given to the user. A complete list of the tasks is found in Appendix B.3.

5.4 Results

The videotaped sessions lasted about three hours and each hour of the videotape took about three hours to evaluate, for a total of 9 hours. For each user, the total time to accomplish each task was measured in seconds and the total number of malfunctions uncovered was measured as well.

5.4.1 Task Completion Time Results

As with the videotaped evaluation of existing tools, in order to avoid bias, the time that involved conversation with the user was subtracted from the total time to accomplish the task. Please refer to Table B-1, which lists the speed and average of task completion time in seconds for each user using BioPathVis.

Figure 5-1 below shows the average task completion times in seconds for both existing tools and BioPathVis (See section 3.1.4.1).

As was mentioned in section 3.1.4.1, tasks that took 400 seconds are those for which the user spent excessive time and eventually gave up. In Figure 5-1, the tasks that are interesting to look at are Tasks 1 to 9 because they allow one to compare the completion times of BioPathVis to other tools. As was mentioned in section 3.1.4.1, tasks 2 and 6 do not show data on MetaCyc, not for the reason that MetaCyc does not support them, but rather because these tasks were accidentally omitted. Task 4 is not supported in MetaCyc and WIT; Task 7 is not supported in BioCarta and ExPASy; Tasks 8 and 9 are not supported in BioCarta, ExPASY, KEGG, and WIT. Tasks 10 to 15 represent tasks that are only available in a certain tool. Out of these, Task 13 is supported by MetaCyc and Task 15 is supported by MetaCyc, ExPASy, and KEGG. However, these tasks were not tested in the videotaped evaluation of existing tools, because of oversight in the early stages of the research. The features that we concentrated on while performing the videotaped evaluation of existing tools had to do mainly with searching. After collecting data from the questionnaires and the brainstorming session, we realized the importance of including some features other than searching in BioPathVis. Unfortunately, data from tools that support these features were not collected. Hence, for some tasks, we cannot compare task completion time for BioPathVis to other tools.

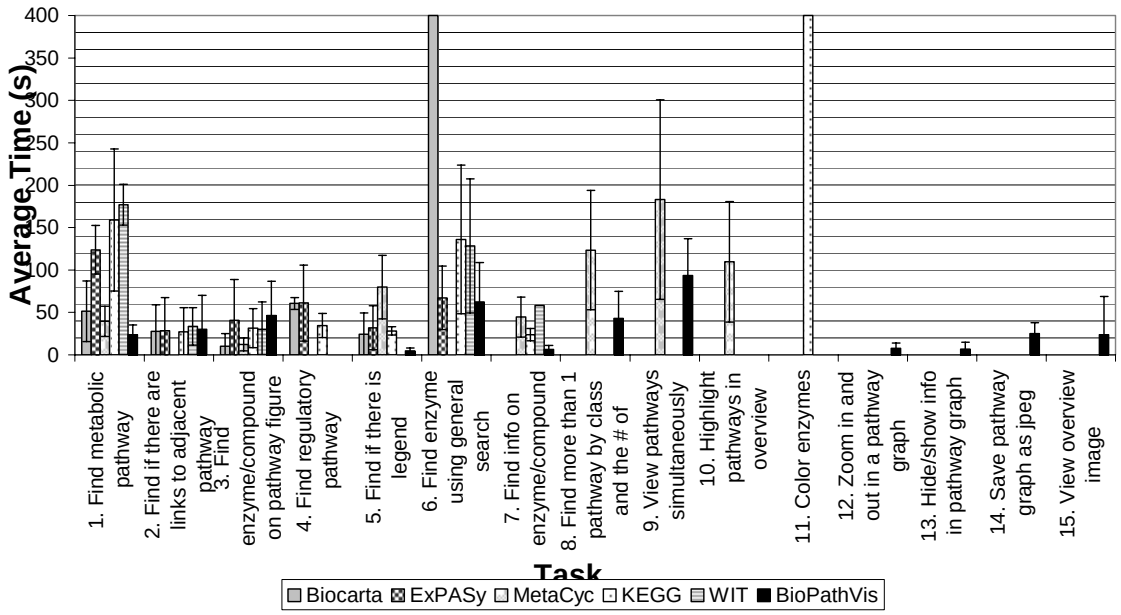


Figure 5-1: Comparing average task completion times for 5 users using BioPathVis and existing tools. The error bars represent the standard deviation

As was done for existing tools, we want to make inferences about the differences among the averages between BioPathVis and existing tools. In order to draw conclusions about whether BioPathVis is better than other tools in accomplishing a certain task, we first conduct ANOVA test to determine if there is a significant difference between the means of the populations, or whether the differences were purely due to random chance. For the tasks for which the difference in the means is not significant, we conduct T-Tests between BioPathVis and each other tools (Please see sections D.1 and D.2 for more details about the ANOVA test and T-Test, respectively).

The results of the ANOVA test are summarized in Table B-4 of Appendix B.4. From the results in the table, we can conclude that for tasks 1, 5, 7, and 8, there is a significant difference between the means of the populations and the difference is not due to random chance.

In order to draw conclusions about which tools are better for tasks 1, 5, 7, and 8, we conducted T-Tests between BioPathVis and each of the other tools. The results are

summarized in Table B-2 of Appendix B.4. From the results in the table, we can conclude that for Task 1 (finding a metabolic pathway), we are 95% confident that if we take a random user from the population and ask the user to accomplish the task using the five tools, the user will accomplish the task faster using BioPathVis (= 23.60 s) than using ExPASy (= 123.75 s), KEGG (= 159.00 s), or WIT (= 177.00 s). For Task 5 (finding if there is a legend), a random user from the population will accomplish the task faster using BioPathVis (= 4.80 s) than using ExPASy (= 32.00 s), KEGG (= 80.00 s), or MetaCyc (= 28.20 s). For Task 7 (finding information on an enzyme/compound), a random user from the population will accomplish the task faster using BioPathVis (= 6.40 s) than using KEGG (= 23.80 s), MetaCyc (= 44.50 s), or WIT (= 58.25 s). For Task 8 (finding more than 1 pathway by class), a random user from the population will accomplish the task faster using BioPathVis (= 42.80 s) than using MetaCyc (= 123.40 s). For the rest of the tasks, we conclude that the observed difference between the two sample means of BioPathVis against other tools is not significant.

5.4.2 Analysis of User Interface Malfunctions

As was done for existing tools, for each malfunction uncovered in BioPathVis, the category the malfunction belongs to, the task the malfunction occurred at, the user number, the problem description, the UI principles violated, the severity level, and recommendations for change were determined (See Table B-10 in Appendix B.5).

Figure 5-2 below shows the number of malfunctions uncovered by the videotaped evaluation of each tool. Note that as was mentioned in section 3.1.4.2, any incorrect behavior of the system, whether it is elated to ‘utility’ of ‘usability’ is recorded as a malfunction.

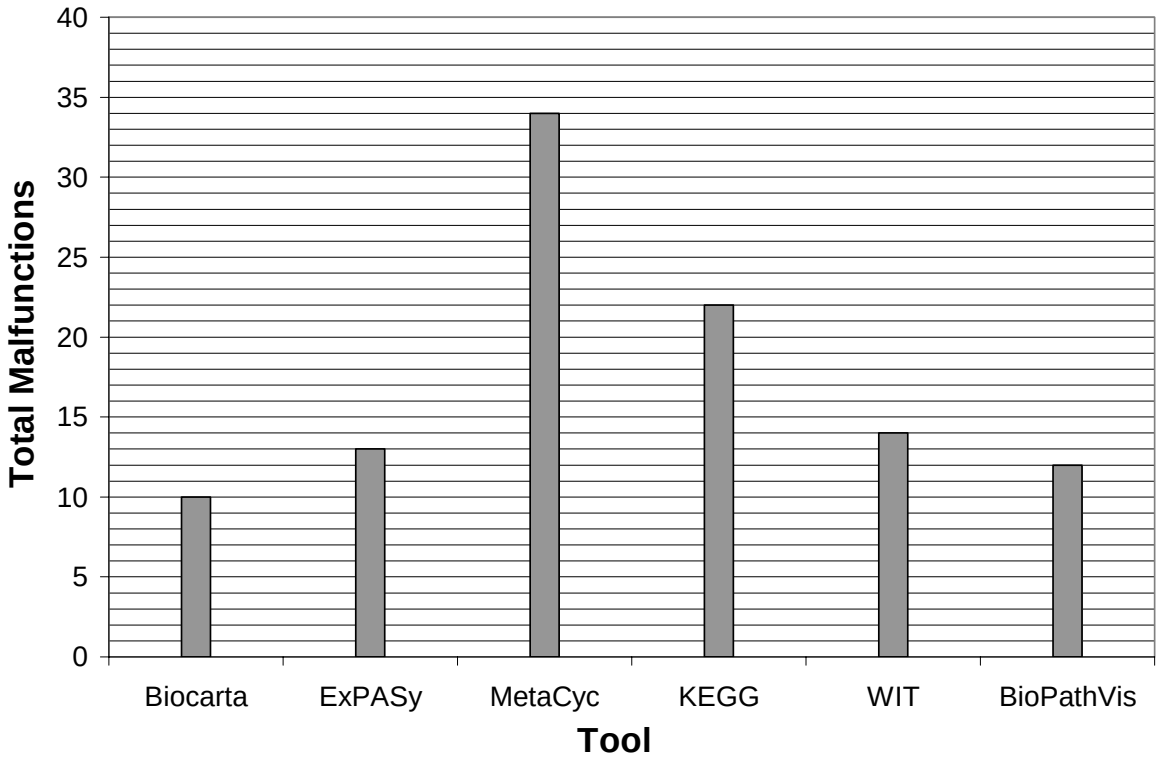


Figure 5-2: Comparing total malfunctions (Data collected from videotaped evaluation with 5 users)

As was discussed in section 3.1.4.2, in some tools, the uncovered malfunctions are related to the additional features that the tools provide. Hence, in order to compare the malfunctions in different tools, we must only consider the tasks that are common between tools. Figure 5-3 below compares the total malfunctions found by all users using BioPathVis and existing tools (See section 3.1.4.2). As the figure shows, four malfunctions in BioPathVis were uncovered in Task 3 (finding compounds on the pathway figure), three in Task 9 (Finding pathways by class), two in Task 15 (viewing an overview image) and one in each of Task 5 (finding a legend), Task 6 (finding a compound), and Task 13 (hiding information on the pathway diagram).

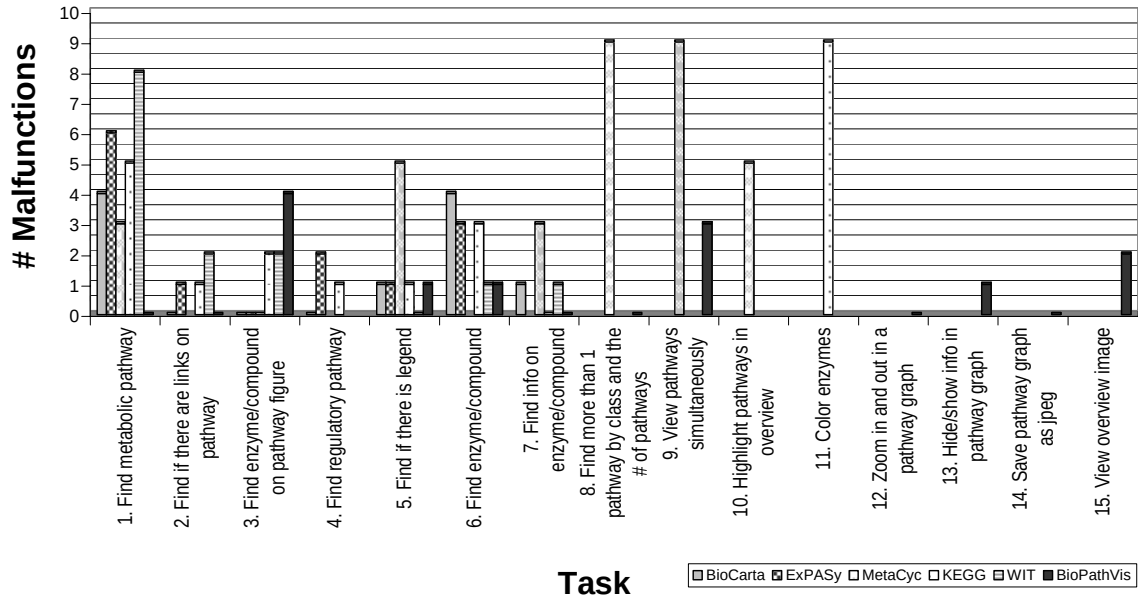


Figure 5-3: Comparing malfunction distribution according to task for BioPathVis and existing tools (Data collected from videotaped evaluation with 5 users)

For finding metabolic pathways (Task 1), finding if there are links on the pathway figure (Task 2), finding information on enzyme or compound (Task 7), finding more than one pathway by class (Task 8), zooming in and out of the pathway graph (Task 12), and saving a pathway graph as a jpeg image (Task 14), BioPathVis had no malfunctions, suggesting that the tool provides a generally intuitive way to perform these tasks.

For finding an enzyme or a compound on the pathway figure (Task 3), BioPathVis had the most malfunctions, as compared to other tools. Two of these malfunctions were due to the use of the "Search in graph" combo box (feature FS19 in Table 4-1), which allows the user to find a compound by selecting it from the combo box that takes the user to the compound and highlights it. As Table 4-1 shows, this feature is only present in BioPathVis. The first malfunction was due to the fact that highlighting is done using a green colored line which is not very noticeable to the user. The second was because one of the users was confused about the "Search in graph" label and the actual combo box; the user thought that the "Search in graph" label is clickable. The remaining two malfunctions were due to scrolling using the mouse middle button being slow and the fact that the legend is incomplete (it does not show

how enzymes or reactions are represented). When comparing BioPathVis malfunctions captured in this task to other tools, we observe that these malfunctions were purely due to providing the additional functionality of doing quick search in the graph. Other tools did not provide such functionality and hence had a lower number of malfunctions. In addition, the severity level (See Table 4-1) of the malfunctions uncovered in BioPathVis while accomplishing this task was low (severity Level 4).

For finding whether there is a legend on the pathway figure (Task 5), BioPathVis had one malfunction. This malfunction was due to the fact that the legend is a floatable toolbar and it is hard to place it in a certain location on the panel to the left of the graph. The severity level (See Table 4-1) of this malfunction is very low (Level 5), since it is a minor cosmetic malfunction.

For finding compounds or enzymes using general search (Task 6), BioPathVis had one malfunction of severity Level 5 (See Table 4-1). This malfunction is due to the fact that selecting a compound type while performing the search does not, by default, select all compounds. This malfunction is more related to performing multiple searches and it is a minor malfunction.

For viewing pathways simultaneously (Task 9), BioPathVis had a lower malfunction count than MetaCyc (three compared to nine). Two of the malfunctions uncovered in BioPathVis have severity level 5 and one had severity level 4. One malfunction involves the text field provided in detailed searching for pathway names, which allows the user to type part of the name and scrolls the list down to the name. This is a feature that only BioPathVis provides. One user was confused about typing the pathway name versus selecting it. Another malfunction occurred when performing multiple selections. Although all users figured out the need to hold down the ‘Ctrl’ or ‘Shift’ keys to perform multiple selections, some users thought that it would be nice to state that somewhere as a reminder. The last malfunction had to do with presenting the search results. The search results are presented in the search panel as a tab next to the quick search tab (See Figure 4-9). Despite the confusion, when users were asked further about this feature and suggestions for improvements, all users commented that

they like how the search results are presented and that they cannot think of a better way to present them.

Malfunctions uncovered in MetaCyc for viewing pathways simultaneously, on the other hand, were greater in number and some had worse severity levels (See Table 4-1). As shown in Table B-8 of Appendix B.5, six of these malfunctions had severity Level 4, one had severity Level 5, one had severity Level 2, and one had severity Level 3. The one with severity Level 2 caused the application to freeze and the one with severity Level 3 caused significant delays to the user. Hence, we can see that BioPathVis presented fewer malfunctions with lower severity that MetaCyc with respect to this task.

For hiding info on the pathway diagram (Task 13), one malfunction of severity Level 4 (See Table 4-1) was uncovered in BioPathVis. This malfunction is due to the fact that the user considered the “View” menu to hide information in the pathway graph, instead of using the tool bar on top of the pathway figure. The “View” menu should have provided that option.

For viewing an overview image (Task 15), BioPathVis had two malfunctions. One malfunction is due to the fact that the user first thought that the user must single click to display a pathway overview. The user later realized that it is necessary to double click. Even though a tool tip was provided, the user never considered it. The second malfunction is due to the fact that the help menu does not provide any help topics. These malfunctions had severity levels 5 and 4, respectively (See Table 4-1).

Figure 5-4 below shows the distribution of the malfunctions found in BioPathVis and existing tools according to the severity levels described in . As we can see in the figure, all of the malfunctions in BioPathVis have severity Level 4 (minor but irritating problems) or Level 5 (minimal error), whereas the other tools had some malfunctions with severity Level 3 (moderate error) and very few in Level 2 (Severe problem).

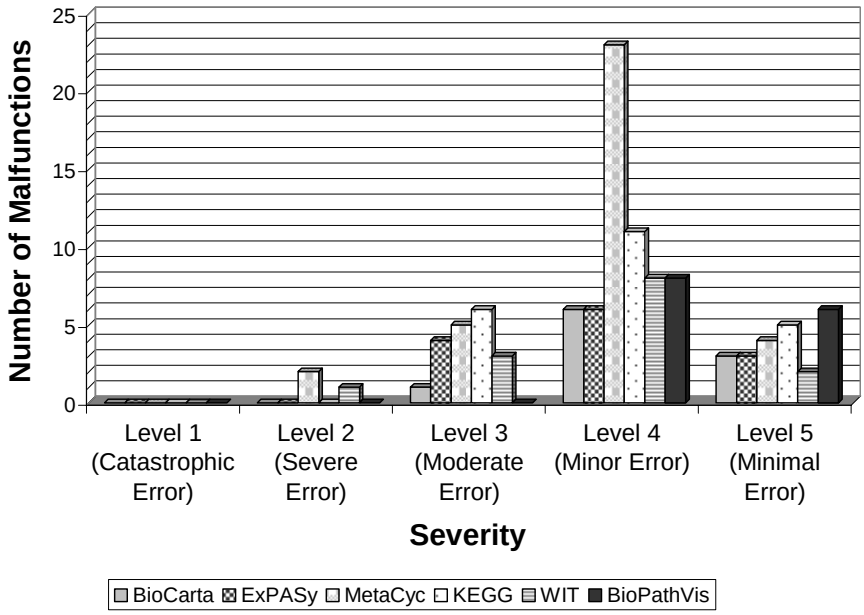


Figure 5-4: Malfunctions distribution according to Severity in BioPathVis and existing tools. (Data collected from videotaped evaluation with 5 users)

In order to account for the oversights in early experimental design, which resulted in excluding MetaCyc from tasks 2, 6, 13, and 15, and ExPASy and KEGG from task 15, we will omit these tasks from the malfunction analysis. Figure 5-5 shows the total malfunctions of BioPathVis and existing tools, omitting those tasks. Note that the same malfunction may appear in many tasks. Hence, if the same malfunction appears in any of tasks 2, 6, 13, or 15, and in another task, then that malfunction is counted because even though it does not appear in tasks 2, 6, 13, or 15, it still appears in another task.

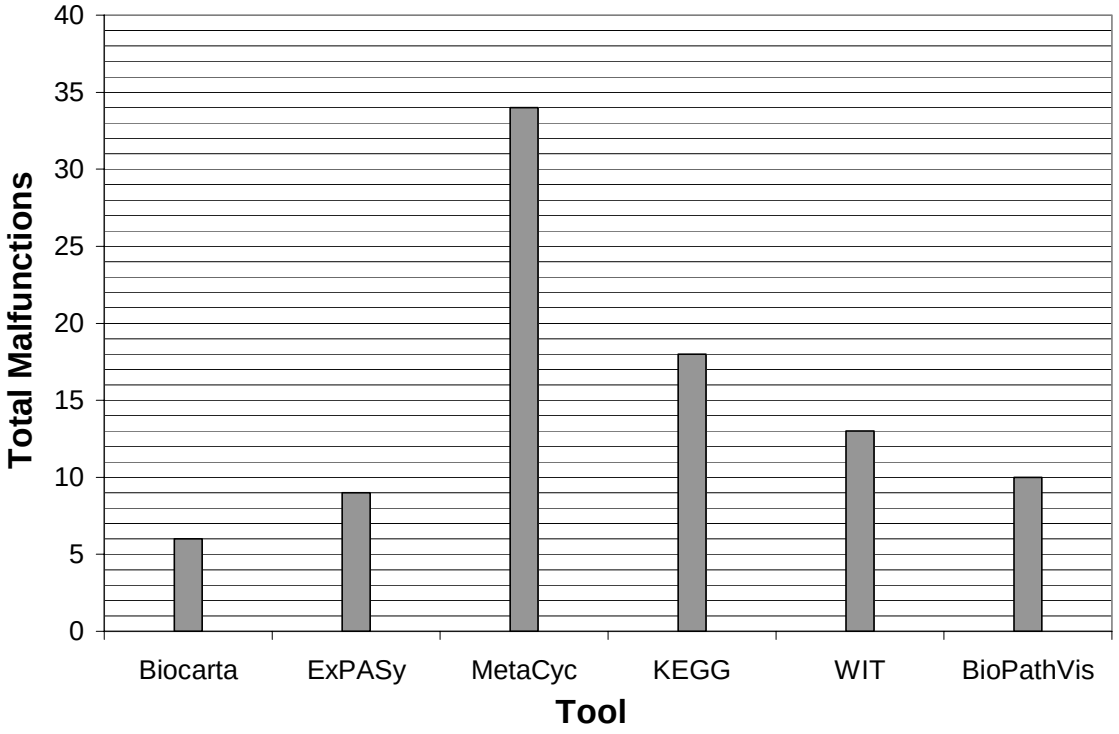


Figure 5-5: Comparing total malfunctions in BioPathVis and existing tools, omitting Tasks 2, 6, 13, and 15 (Data collected from videotaped evaluation with 5 users)

Figure 5-5 shows that BioCarta has the lowest number of malfunctions, followed by ExPASy, followed by BioPathVis, followed by WIT, followed by KEGG, and followed by MetaCyc. Despite the fact that BioCarta has the lowest number of malfunctions, it has a low feature set number (8 out of 51) and a low total weight (12.5 out of 78.5 or 15.9%), according to Table 4-1. ExPASy also has a low feature set number (7 out of 51) and a low total weight (17.6 out of 78.5 or 22.4%). WIT has the lowest feature set number (4 out of 51) and a lowest total weight (7.8 out of 78.5 or 9.9%). KEGG has a higher feature set number than BioCarta, ExPASy, and WIT (14 out of 51) and a higher total weight (24.9 out of 78.5 or 31.7%). MetaCyc has yet a higher feature set number than the above mentioned tools (24 out of 51) and a higher total weight (38.0 out of 78.5 or 48.4%). BioPathVis, on the other hand, has the highest feature set number (28 out of 46) and total weight (52.1 out of 78.5 or 66.4%) as compared with other tools. This indicates that BioPathVis tested for more functionality that were more relevant to the visualization of biochemical pathways.

Figure 5-6 below shows the malfunction distribution according to severity, omitting tasks 2, 6, 13, and 15. Again, since the same malfunction may appear in many tasks, some malfunctions were not omitted even though they appeared in tasks 2, 6, 13, and 15.

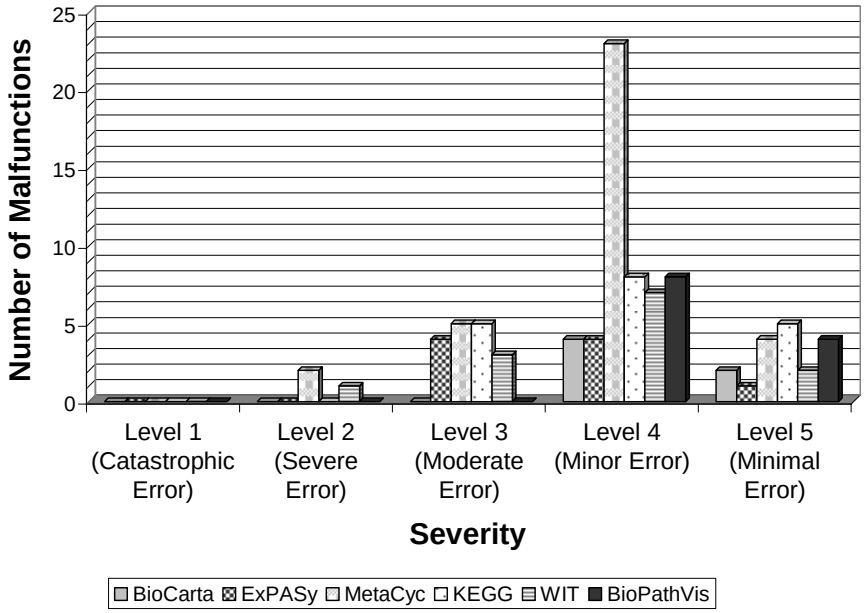


Figure 5-6: Malfunctions distribution according to severity in BioPathVis and existing tools, omitting Tasks 2, 6, 13, and 15 (Data collected from videotaped evaluation with 5 users)

We can draw the same conclusions as compared to Figure 5-4: That all malfunctions in BioPathVis were in Level 4 (minor but irritating problems) or Level 5 (minimal error), whereas the other tools had some malfunctions in Levels 3 (moderate error) and very few in Level 2 (Severe problem).

5.4.3 Follow up Questions Results

In order to gather more data on user preferences, the same users who conducted the videotaped evaluation were asked to fill out a follow up questionnaire to get their feedback on BioPathVis (See Appendix B.6). The questionnaire consisted of two sections. The first

was used to collect user preference data, in terms of the ease of use of features of BioPathVis, problems they faced, suggestions for improvements, how much they liked BioPathVis, and how likely they were to use it. The last section consisted of personal data about the users, such as their age, profession, education, background, and experience with biochemical pathways and computers.

Figure 5-7 below shows the ease of use of features of BioPathVis. Users were asked to rate the ease of use of BioPathVis using five discrete values, which are "Very easy", "Easy", "Moderately easy", "Difficult", and "Very difficult". Then, data was collected from all users and the five discrete values were converted to numbers from 1 to 5, where 1 is "Very difficult" and 5 is "Very easy". The average ease of use is calculated using the following formula:

Average Ease of Use

$$= \frac{(\#\,\text{Very difficult}) * 1 + (\#\,\text{Difficult}) * 2 + (\#\,\text{Moderately easy}) * 3 + (\#\,\text{Easy}) * 4 + (\#\,\text{Very easy}) * 5}{5}$$

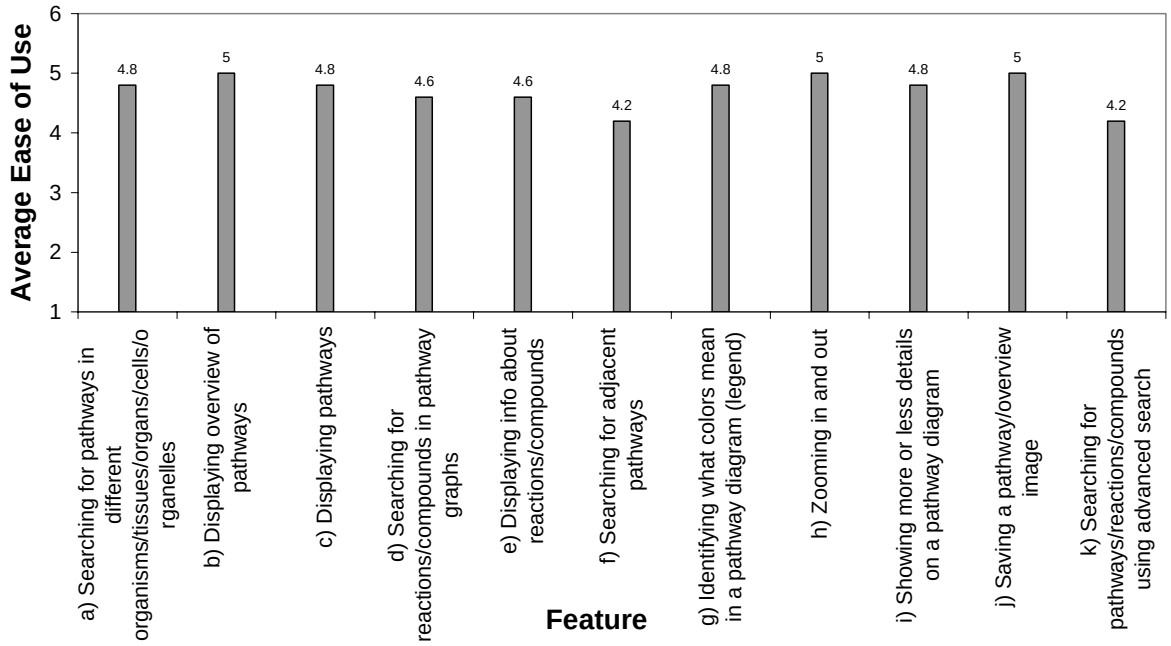


Figure 5-7: Average ease of use of features of BioPathVis on scale from 1 to 5, where 1 is "Very difficult" and 5 is "Very easy" (Data collected from 5 users after videotaped evaluation)

Figure 5-7 shows that the average ease of use of most features of BioPathVis was close to 5 (Very easy). Searching for adjacent pathways and searching for pathways or reactions or compounds using advanced search received an average ease of use value that is close to 4 (Easy). The reason why advanced searching was a little more difficult for some users is because users were a bit confused about the search results being displayed as a tab containing a tree view of search results. Other minor malfunctions related to advanced searching can be found in Table B-10 in Appendix B.5.

Users were asked to rate how much they liked or disliked BioPathVis as a biochemical pathway visualization tool, using five discrete values, which are "Very much like it", "Like it", "Neutral", "Dislike it", and "Very much dislike it". Three users said that they "Very much like it" and two said that they "Like it". The average is calculated using the following formula:

Average Rating of BioPathVis

$$= \frac{(\#\text{Very much like it}) * 1 + (\#\text{Like it}) * 2 + (\#\text{Neutral}) * 3 + (\#\text{Dislike it}) * 4 + (\#\text{Very much dislike it}) * 5}{5}$$

The average is 1.4, which is close to the "Very much like it" rating.

Users were also asked to rate how likely they are to use BioPathVis as a biochemical pathway visualization tool, using five discrete values, which are "Very likely", "Likely", "Neutral", "Unlikely", and "Very unlikely". Three users said that they are "Very likely" to use BioPathVis and two said they are "Likely" to use it. The average is calculated using the following formula:

Average Likelihood of using BioPathVis

$$= \frac{(\#\text{Very likely}) * 1 + (\#\text{Likely}) * 2 + (\#\text{Neutral}) * 3 + (\#\text{Unlikely}) * 4 + (\#\text{Very unlikely}) * 5}{5}$$

The average is 1.4, which is close to the "Very likely" rating.

When asked about suggestions for improvements, users made the following recommendations:

1. For zooming, a quick zooming is nice to have. At the very least, a full page view. Also a percentage view would be nice to have (e.g. 100%, full page, etc – like in MS word).
2. Change the adjacent pathway color to a color like purple instead of yellow, because the color of Nucleotides is orange, which is close to yellow.
3. Add more information to the "Compound Info" dialog box.
4. Add more information to the legend.
5. Use single clicking as opposed to double clicking to view biochemical pathways or information about reactions and compounds.

All of the above recommendations will be applied. However, special attention must be applied to the last recommendation. For time consuming tasks, such as displaying biochemical pathways, double clicking is more appropriate, since single clicking would be time consuming. One possibility is to allow single clicking, but provide a 'Stop' button, similar to the one present in web browsers.

When asked to make more comments about BioPathVis, the two users, who also did the videotaped evaluation of existing tools, commented that they really liked BioPathVis. They liked the use of colors and its ease of use. One user said that it's "Definitely a lot easier than other ones" and that "If your goal is to make it more user friendly, then you met the target". The user who studied Bio-pharmacy commented that "It's really good. I have to admit", "It's clear", "Really well done", "Great interface", and "It took me nothing to learn how to use it".

5.4.4 Future Enhancements

The videotaped evaluation results showed that some enhancements must be made to BioPathVis. These are summarized both in Table B-10 in Appendix B.5 and at the end of section 5.4.3. In addition, more functionality should be added to take into account the rest of the use cases described in section 4.1 and the features described in Table 4-1.

Another enhancement that must be made is to ensure database independence. The DbInterface package could be extended to allow support for databases other than Oracle. But most importantly, to add full support for reading data from XML files. Although the latter was implemented during the first stages of the design, it still requires some enhancements and modifications. Plus data needs to be populated in XML files, by importing it from our current Oracle database.

Furthermore, some additional features can be added. For instance, adding E.C. numbers for enzymes, adding regulatory pathways, and adding activators and inhibitors on pathway diagrams. The importance of these features is described in Table 3-4.

With respect to adding customizable preferences (See use case 46 in section 4.1), users should be able to change the representation used in biochemical pathways (e.g., the use of the enzyme's E.C. number as opposed to the name in the pathway figure). In addition, users should be able to customize the pathway graph to show or hide certain information upon displaying the pathway graph. It should also allow the user to show or hide components of the user interface, such as the legend, or status bar, upon application startup.

Another enhancement that could be made to BioPathVis is allowing adjacent pathway nodes to expand to show more than 1 pathway in the same pathway graph. As was mentioned in section 3.3.2, users commented that there should be a limit of three on how many pathways can be displayed in one graph. Users should be given the capability to change that number in the preferences menu.

Since BioPathVis is a visualization tool, it can form the basic module that can be extended with other modules. One example of such extension modules is an editor as was mentioned earlier. Another is a simulation module, which can be qualitative or quantitative in nature.

Chapter 6 Concluding Remarks

The objective of this research has been the design of an easy to use, intuitive, and comprehensive user interface for biochemical pathway visualization. To accomplish this task, we first reviewed previous work on the usability of information visualization, in particular biochemical pathway visualization, and then we conducted three different kinds of user studies to gather information about some of the existing biochemical pathway visualization tools, namely BioCarta, ExPASy, KEGG, MetaCyc, and WIT. The three user studies are (1) conducting videotaped evaluation sessions of existing biochemical visualization tools, (2) collecting questionnaires, and (3) conducting a brainstorming session. The results of these studies, which are summarized in Chapter 3, were used to define the requirements and then develop a new biochemical pathway visualization tool, BioPathVis. The shortcomings of the existing tools, which were uncovered by the videotaped evaluation, allowed us to understand how to build a more effective and intuitive tool. In addition, the results collected from the questionnaires and the brainstorming session allowed us to define the difficulties biochemists encounter when visualizing biochemical pathways, as well as the most important features that biochemists would like to have in a biochemical pathway visualization tool.

The results described in Chapter 4 have shown that we have met our objective. In particular, BioPathVis has the highest feature set count (28 out of 41), followed by MetaCyc (24), followed by KEGG (14), followed by BioCarta (8), followed ExPASy (7), and followed by WIT (4). In addition, BioPathVis has the highest total feature set weight (121.1 out of 178.8 or 67.7%), followed by MetaCyc (103.1 or 57.7%), followed by KEGG (62.7 or 35.1%), followed by ExPASy (30.6 or 17.1%), followed by BioCarta (37.5 or 21.0%), and followed by WIT (18.3 or 10.2%) (See Table 4-1 and discussion in section 5.4.2). This indicates that BioPathVis contains functionality that is more relevant to the visualization of biochemical pathways

In addition, the videotaped evaluation of BioPathVis showed that for Task 1 (finding a metabolic pathway), Task 5 (finding if there is a legend), Task 7 (finding information on an enzyme/compound), and Task 8 (finding more than 1 pathway by class), we are 95% confident that BioPathVis showed superiority compared to some of the other tools that provide the same functionality (See section 5.4.1).

Furthermore, for most of the tasks (e.g. finding metabolic pathways (Task 1), finding if there are links on the pathway figure (Task 2), finding info on enzyme compound (Task 7), finding more than 1 pathway by class (Task 8), zooming in and out of the pathway graph (Task 12), and saving a pathway graph as a jpeg image (Task 14)), BioPathVis had no malfunctions, suggesting that the tool makes these tasks intuitive (See Figure 5-3 and discussion in section 5.4.2). For the rest of the tasks, some of the malfunctions in BioPathVis were due to the additional features that BioPathVis provides, such as quick searching for compounds in a pathway graph. In general, malfunctions in BioPathVis had lower severity levels as compared with other tools (See Figure 5-4, Figure 5-6, and discussion in section 5.4.2).

Finally, the results in Figure 5-7 and section 5.4.3 show that the average ease of use of most features of BioPathVis is close to 5 (Very easy) and some close to 4 (Easy). Also, the average rating of how much users liked BioPathVis is 1.4, which is closer to the "Very much like it" rating. The average likelihood of using BioPathVis as a biochemical pathway visualization tool is 1.4, which is closer to the "Very likely" rating. In addition, user comments on BioPathVis described in section 5.4.3 are very positive, indicating its success as a biochemical pathway visualization tool.

There are many reasons that BioPathVis presented an improved UI as compared with other tools. One reason is following UI guidelines while designing the tool. Another reason has to do with the methodology of collecting information on how to design BioPathVis, which is shown to be valid and effective. Studying existing tools allowed us to combine the best features and utilities of these tools in BioPathVis, as well as to avoid practices that caused confusion to the user. In addition, conducting user experiments and reviewing literature allowed us to gain a better understanding of which features are more desirable. Carefully

choosing between alternative design decisions and involving the user in making such decisions is yet another reason behind the improved UI design of BioPathVis.

Now that we have discussed the improvements shown by the BioPathVis tool, we would like to make some generalizations to other systems. Subsequently, we will discuss some research limitations that have led to some valuable missing data. At the end, we will discuss some interesting areas for further research.

6.1 Application to Other Systems

Since biochemical pathway visualization is a special case of information visualization, we will attempt to make some generalizations to the design of user interfaces involved in information visualization.

The first general suggestion that arises from this research is to consider the importance of providing navigation and control capabilities directly in the area of focus of users. This idea has been discussed in section 2.3.4, for searching. However, the videotaped evaluation results show that this idea should be applied everywhere. For instance, when asked to hide a legend, users tried clicking on the legend (See Table B-10). Users only considered the "View" menu or toolbar later. Users preferred having the capability to hide the legend right on the legend, as opposed to having a floatable window (See Table B-10). In general, users tend to look at the main graph when performing any task that is related to the graph, before consulting the menu.

Another general suggestion is to consider the importance of choice of words, as well as their consistency, when describing actions. For instance, users consulted the "View" menu when asked to hide or show information (See Table B-10). The "View" menu should have provided that capability. Carefully choosing words is a well-known UI guideline; our experiences in this thesis show that it needs to be re-iterated.

A third idea that can be applied to other systems relates to the detailed search functionality. The detailed search dialog box contains a text field that allows the user to type in part of the pathway name; this in turn takes the user to the closest matching pathway name. The user can select multiple pathways by holding down the 'Ctrl' or 'Shift' keys while making the selections. Although this idea of combining of matching keywords and selections is not common in standard applications, all users were capable of figuring it out with relative ease, suggesting that the use of such features would be intuitive in other systems too.

A final generalizable result has to do with displaying results as a tab in the quick search area. Although some users were confused by this feature, when they were asked further about this feature and suggestions for improvements, all users commented that they liked how the search results are presented and that they cannot think of a better way to present them (See section 5.4.2).

6.2 Research Limitations

This section describes some research limitations that might limit the generalizability of this work.

The most important class of limitations arises from the fact that there is inherent subjectivity in several aspects of the research. In the questionnaires, the end-users gave subjective responses, but more importantly, evaluation of what constituted a malfunction and the level of importance of that malfunction was determined solely by the researcher. Some of the evidence of usability in this thesis should therefore be taken as suggestive rather than definitive. In an industrial setting, expensive and independent evaluations would normally be employed, and several people could independently perform the evaluation. This was not feasible in the context of a Masters thesis.

A second class of limitations arises from the relatively small number of users employed. More users would have resulted in greater confidence in the data. However, we believe that

the number of users was adequate to validate that the system was basically usable, and to find most of the important malfunctions.

Another class of limitations relates to the fact that some of the existing pathway tools have changed. At the time the videotaped evaluation of existing tools was performed, MetaCyc looked quite different from the way it was at the time the thesis was being completed. In fact, many of the modifications to MetaCyc were due to a report on usability that we sent to the developers of MetaCyc suggesting enhancements to the tool. In addition, WIT does not exist anymore, and it has been replaced by the SEED and PUMA2 [40]. The website of KEGG has also changed [20]. This limitation is inherent in research on moving targets; however, the fact that data from the early stages of our research was used to improve other tools, validates our methodology and shows that the research provided a valuable contribution.

Additional limitations in this research have to do with oversights in the early stages of experimental design. For instance, in the videotaped evaluation, as was mentioned in sections 3.1.4.1, 3.1.4.2, 5.4.1, and 5.4.2, some tasks were not performed for certain tools, not because the tools did not provide the functionality, but rather because the researcher did not take these features into account. These are tasks 2, 6, 13 (for MetaCyc), and 15 (for ExPASy, KEGG, and MetaCyc). Due to the inherent nature of human-centred research, it was not feasible to go back and ask the users to perform the tasks that had been missed; hence no comparisons can be made for these tools with other tools.

In addition, results on important features collected from the questionnaire and brainstorming session did not allow us to compare these features using the same scale. Basically, Figure 3-9 and Table 3-3 in section 3.2.2 allowed us to assign an average importance value to each feature, whereas Table 3-5 and Table 3-6 in section 3.3.2 did not allow us to make such an assignment. Therefore, data on the average importance of features from these tables were not included in the feature set described in Table 4-1.

Finally, a follow-up questionnaire was not performed for the existing tools. Hence, data on how much users liked existing tools and how likely they are to use them is not available to

compare it to BioPathVis.

Despite the above limitations, we were still capable of analyzing the results, and making some important conclusions.

6.3 Areas for Further Research

There are many interesting areas that we came across while conducting this research, for which we were not able to investigate further because of time constraints. The first is the 3D representation of biochemical pathways. As was discussed in section 1.5, not much work has been conducted on building a tool to visualize biochemical pathways in 3D and to study the usability of such representation. In fact, some users in the brainstorming session (See section 3.3.2) said that it would be an interesting topic to investigate and suggested the use of depth to organize biochemical pathways in layers.

Moreover, further research should be done on the usability of other features that have not yet been implemented in BioPathVis. One feature involves incorporating regulatory pathways and regulatory control into the biochemical pathway representation tool. Regulatory pathways have different representations from metabolic pathways, and hence more research should be done on how to represent them. Another feature is the use of animation, both in tutorials on how to use the tool and to describe interactions between different pathways.

Another feature is incorporating editing capabilities in BioPathVis, which would allow the user to edit information dynamically using the tool. The editor should provide controls to allow the user to add or edit hierarchical information (e.g. organisms, organs, etc.), as well as reactions, compounds, biochemical pathways (including both metabolic and regulatory pathways), adjacent pathway links, and information about the different reactions and compounds. Studying the usability of the editor is yet another area that could be further investigated.

References

[1] Auber, D. (2002). "Tulip: an information visualization software for huge graphs", *IEEE Computer Graphics Forum (in revision)*.

[2] Baldonado, M., Woodruff, A., and Kuchinsky, A. (2000). "Guidelines for Using Multiple Views in Information Visualization", *Proc. AVI*, Palermo, Italy.

[3] Becker, M. Y. and Rojas, I. (2001). "A Graph Layout Algorithm for Drawing Metabolic Pathways", *Bioinformatics*, 17(5): 461-467.

[4] BioCarta. (2000). [www.biocarta.com], website access date: 2003.

[5] Chen, C. and Czerwinski, M. (2000). "Empirical Evaluation of Information Visualizations: An Introduction", *International Journal of Human Computer Studies*, 53: 631 – 635.

[6] Chen, C and Czerwinski, M. (1997). "Spatial Ability and Visual Navigation: An Empirical Study", *New Reviews of Hypermedia and Multimedia,* 3, 67-89.

[7] Delbecq A. L., Van de Ven, A. H., Gustafson, D. H. (1975). "Group Techniques for Program Planning". *Scott, Foresman & Co*, Glenview, IL.

[8] Deville, Y., Gilbert, D., Helden, J., and Wodak, S. (2003). "An overview of data models for the analysis of biochemical pathways", *Briefings in Bioinformatics*, **4**(3), 246–259, (2003)

[9] Entelos Inc. [www.entelos.com], Foster City, California, website access date: 2003.

[10] ExPASy. (2003) "Biochemical Pathways", Boehringer Mannheim GmbH. [www.expasy.org/cgi-bin/search-biochem-index] , website access date: 2003.

[11] Fowler, M., Scott, K. (1997). "UML Distilled: Applying the Standard Object Modeling Language". *Addison Wesley Longman, Inc*.

[12] Gene Network Sciences. [www.gnsbiotech.com], Ithaca, New York, website access date: 2003.

[13] Herman, I., Melançon, G., and Marshall, S. (2000, January-March). "Graph Visualization and Navigation in Information Visualization: A Survey", *24 IEEE Transactions on Visualization and Computer Graphics,* vol. 6, no. 1.

[14] Institute für Zell- und Organsimulation GmbH. [www.ifcos.com/], website access date: 2003.

[15] JGraph Ltd. (2004) [www.jgraph.com], website access date: 2003.

[16] Jr, E., Grechkin, Y., Mikhailova, N., and Selkov, E. (1998). "MPW: The Metabolic Pathways Database", *Nucleic Acids Research*, 26 (1): 43-45.

[17] Kanehisa, M., Goto, S., Ogata, H., Sato, K., Fujibuchi, W., and Bono, H. (2000). "KEGG: Kyoto Encyclopedia of Genes and Genomes", *Nucleic Acids Res*: 28, 27-30.

[18] Karp, P. (2001). "Pathway Databases: A Case Study in Computational Symbolic Theories", *Science* 293: 2040-2044.

[19] Karp, P. and Paley, S. (1996). "Integrated Access to Metabolic and Genomic Data", *Journal of Computational Biology*, 3(1): 191-212.

[20] KEGG. (2003). [www.kegg.com], website access date: 2003.

[21] Lethbridge, Timothy. (2002). "CSI 5122 Software Usability Class Notes: Unit E: Design Guidelines".

[22] Lethbridge, Timothy. (2001). "SEG 3120 User Interface Design and Implementation".

[23] Lethbridge, T.C. and Skuce, D. (1992, October), "Integrating Techniques for Conceptual Modeling", *Proc. 7th Knowledge Acquisition for Knowledge-based Systems Workshop,* Banff, Alberta, pp. 15.1-15.20. [www.site.uottawa.ca/~tcl/papers/km/IntegratingTechniques.html], website access date: 2003.

[24] Luzzardi, P., Winckler, M., Cava, R., Pimenta, M., Nedel, L., and Freitas, D. (2002). "Usability of Information Visualization Techniques", *Proc. AVI,* Trente, Italy, 22-24.

[25] Mendenhall, W., Beaver, R., Beaver, B. (1999). "Introduction to Probability and Statistics". Brooks/Cole Publishing Company.

[26] MetaCyc. (2003). [www.biocyc.com], website access date: 2003.

[27] Michal, G. Biochemical Pathways (poster). Boehringer Mannheim GmbH (1993).

[28] Michal, G. (1999). "Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology". *John Wiley & Sons*, New York.

[29] Miller, I. And Freund, J. (1977). "Probability and Statistics for Engineers". *Prentice Hall*. New Jersey.

[30] Moran, L., Scrimgeour, K., Horton, H., Ochs, R., and Rawn, D. (1994). "Biochemistry (Second Edition)". *Neil Patterson Publishers Prentice Hall*, New Jersey.

[31] Neilson, J. (2001). "2D is better than 3D." [www.useit.com], website access date: 2003.

[32] Neilson, J. (2000). "Is Navigation Useful?" [www.useit.com], website access date: 2003.

[33] Overbeek, R., Larsen, N., Pusch, G., D'Souza, M., Jr, E., Kyrpides, N., Fonstein, M., Maltsev, N., and Selkov, E. (2000). "WIT: Integrated System for High-throughput Genome Sequence Analysis and Metabolic Reconstruction", *Nucleic Acids Res*, 28:123-125.

[34] Parker, G., Frank, G., and Ware, C. (1997). "Visualization of Large Nested Graphs in 3D: Navigation and Interaction", *Journal of Visual Languages and Computing*, 9 (5): 299-317.

[35] Physiome Sciences Inc. [www.physiome.com]. Princeton, New Jersey, website access date: 2003.

[36] Rzhetsky, A., Koike, T., Kalachikov, S., Kra, P., Yu, H., and Friedman, C. (1999). "A Knowledge Model for Analysis and Simulation of Regulatory Networks in Bioinformatics Studies Aiming at Disease Gene Discovery", *AMIA'99 Poster Presentations*.

[37] Ware, C., Frank, G., Parkhi, M., and Dudley, T. (1997). "Layout for Visualizing Large Software Structures in 3D", *Proc. Visual '97,* San Diego, 215-223.

[38] Ware, C. and Plumlee, M. (2002, May). "Zooming, Multiple Windows, and Visual Working Memory", *Proc. AVI,* Trente, Italy, 22-24.

[39] Wilson, C. (2001). "Usability Interface", [www.stcsig.org/usability/newsletter/9904-severity-scale.html], website access date: 2003.

[40] WIT. (2000). [http://wit.mcs.anl.gov/WIT2/]], website access date: 2003.

[41] yWorks. (2000) "Products: Extension packages – yWays". [http://www.yworks.com/en/products_yfiles_extensionpackages_ep2.htm], website access date: 2005.

Appendix A Classification of Biochemical Pathways

Table A-1: A list of known metabolic pathways, classified by molecule metabolized, as suggested by KEGG (Paraphrased from [20])

Metabolic Pathway Class	Metabolic Pathway Name
Carbohydrate Metabolism	Glycolysis / Gluconeogenesis
	Citrate Cycle (TCA Cycle)
	Pentose Phosphate Cycle
	Pentose and Glucuronate Interconversions
	Fructose and Mannose Metabolism
	Galactose Metabolism
	Ascorbate and Aldarate Metabolism
	Pyruvate Metabolism
	Glyoxylate and Dicarboxylate Metabolism
	Propanoate Metabolism
	Butanoate Metabolism
	C5-Branched Dibasic Acid Metabolism
	Inositol Metabolism
Energy Metabolism	Oxidative Phosphorylation
	ATP Synthesis
	Photosynthesis
	Carbon Fixation
	Reductive Carboxylate Cycle (CO_2 Fixation)
	Methane Metabolism
	Nitrogen Metabolism
	Sulfur Metabolism
Lipid Metabolism	Fatty Acid Biosynthesis
	Fatty Acid Metabolism
	Synthesis and Degradation of Ketone Bodies
	Sterol Biosynthesis

Metabolic Pathway Class	Metabolic Pathway Name
Lipid Metabolism	Bile Acid Biosynthesis
	C21-Steroid Hormone Metabolism
	Androgen and Estrogen Metabolism
Nucleotide Metabolism	Purine Metabolism
	Pyrimidine Metabolism
	Nucleotide Sugars Metabolism
Amino Acid Metabolism	Glutamate Metabolism
	Alanine and Aspartate Metabolism
	Glycine, Serine and Threonine Metabolism
	Methionine Metabolism
	Cysteine Metabolism
	Valine, Leucine, and Isoleucine Degradation
	Valine, Leucine and Isoleucine Biosynthesis
	Lysine Biosynthesis
	Lysine Degradation
	Arginine and Proline Metabolism
	Histidine Metabolism
	Tyrosine Metabolism
	Phenylalanine Metabolism
	Tryptophan Metabolism
	Phenylalanine, Tyrosine and Tryptophan Biosynthesis
	Urea Cycle and Metabolism of Amino Groups
Metabolism of Other Amino Acids	Beta-Alanine Metabolism
	Taurine and Hypotaurine Metabolism
	Aminophosphonate Metabolism
	Selenoamino Acid Metabolism
	Cyanoamino Acid Metabolism
	D-Glutamine and D-Glutamate Metabolism
	D-Arginine and D-Ornithine Metabolism
	D-Alanine Metabolism
	Glutathione Metabolism

Metabolic Pathway Class	Metabolic Pathway Name
Metabolism of Complex Carbohydrates	Starch and Sucrose Metabolism
	N-Glycans Biosynthesis
	O-Glycans Biosynthesis
	N-Glycans Degradation
	Aminosugars Metabolism
	Lipopolysaccharide Biosynthesis
	Peptidoglycan Biosynthesis
	Glycosaminoglycan Degradation
	Chondroitin / Heparan Sulfate Biosynthesis
	Keratan Sulfate Biosynthesis
Metabolism of Complex Lipids	Glycerolipid Metabolism
	Glycosylphosphatidylinositol(GPI)-Anchor Biosynthesis
	Inositol Phosphate Metabolism
	Sphingophospholipid Biosynthesis
	Phospholipid Degradation
	Sphingoglycolipid Metabolism
	Blood Group Glycolipid Biosynthesis - Lact Series
	Blood Group Glycolipid Biosynthesis – Neolact Series
	Globoside Metabolism
	Ganglioside Biosynthesis
	Prostaglandin and Leukotriene Metabolism
Metabolism of Cofactors and Vitamins	Thiamine Metabolism
	Riboflavin Metabolism
	Vitamin B6 Metabolism
	Nicotinate and Nicotinamide Metabolism
	Pantothenate and CoA Biosynthesis
	Biotin Metabolism
	Folate Biosynthesis
	One Carbon Pool by Folate
	Retinol Metabolism
	Porphyrin and Chlorophyll Metabolism
	Ubiquinone Biosynthesis

Metabolic Pathway Class	Metabolic Pathway Name
Biosynthesis of Secondary Metabolites	Terpenoid Biosynthesis
	Flavonoids, Stilbene and Lignin Biosynthesis
	Alkaloid Biosynthesis I
	Alkaloid Biosynthesis II
	Penicillins and Cephalosporins Biosynthesis
	β-Lactam Resistance
	Streptomycin Biosynthesis
	Erythromycin Biosynthesis
	Tetracycline Biosynthesis
	Clavulanic Acid Biosynthesis
	Puromycin Biosynthesis
Biodegradation of Xenobiotics	Caprolactam Degradation
	Biphenyl Degradation
	Toluene and Xylene Degradation
	γ-Hexachlorocyclohexane Degradation
	3-Chloroacrylic Acid Degradation
	1,1,1-Trichloro-2,2-Bis(4-Chlorophenyl)Ethane (DDT) Degradation
	2,4-Dichlorobenzoate Degradation
	1,2-Dichloroethane Degradation
	Tetrachloroethene Degradation
	Styrene Degradation
	1,4-Dichlorobenzene Degradation
	Nitrobenzene Degradation
	Ethylbenzene Degradation
	Fluorene Degradation
	Carbazole Degradation
	Benzoate Degradation via CoA Ligation
	Benzoate Degradation via Hydroxylation
	Atrazine Degradation

Table A-2: A list of regulatory pathways, classified by biological processes, as suggested by KEGG (Paraphrased from [20])

Biological Process	Metabolic Pathway Class	Metabolic Pathway Name
Genetic Information Processing	*Transcription*	RNA Polymerase
		Basal Transcription Factors
		mRNA Biosynthesis (Bacteria Eukaryotes)
		tRNA Biosynthesis (Bacteria Eukaryotes)
		rRNA Biosynthesis (Bacteria Eukaryotes)
	Translation	Ribosome
		Translation Factors
		Aminoacyl-tRNA Biosynthesis
		Protein Biosynthesis (Bacteria, Eukaryotes)
	Sorting and Degradation	Protein Export
		Type II Secretion System
		Type III Secretion System
		Type IV Secretion System
		Ubiquitin Mediated Proteolysis
		Proteasome
	Replication and Repair	DNA Polymerase
Environmental Information Processing	*Membrane Transport*	ABC Transporters, Prokaryotic Type
		ABC Transporters, ABC-2 and Other Types
		Phosphotransferase System (PTS)
	Signal Transduction	Two-component System
		MAPK Signaling Pathway
		Second Messenger Signaling Pathway
		Phosphatidylinositol Signaling System
	Ligand-Receptor Interaction	G Protein Coupled Receptors
		Ion Channel Receptors
		Cytokine Receptors

Biological Process	**Metabolic Pathway Class**	**Metabolic Pathway Name**
Cellular Processes	*Cell Motility*	Bacterial Chemotaxis
		Flagellar Assembly
	Cell Growth and Death	Cell Cycle
		Apoptosis
	Cell Communication	Integrin-Mediated Cell Adhesion
		Cadherin-Mediated Cell Adhesion
	Development	Wnt Signaling Pathway
		Notch Signaling Pathway
		Dorso-Ventral Axis Formation
	Behavior	Circadian Rhythm
Human Diseases	*Neurodegenerative Disorders*	Alzheimer's Disease
		Parkinson's Disease
		Amyotrophic Lateral Sclerosis (ALS)
		Huntington's Disease
		Dentatorubropallidoluysian Atrophy (DRPLA)
		Prion Disease

Appendix B Videotaped Evaluation

B.1 Informed Consent Form

The informed consent form, approved by the University of Ottawa's Research Ethics Board, and used in this study, will be placed on the web site of this thesis: http://www.site.uottawa.ca/~tcl/gradtheses/rkhartab.

B.2 Tutorial

The tutorial given prior to the videotaped evaluation will be placed on the web site of this thesis: http://www.site.uottawa.ca/~tcl/gradtheses/rkhartab.

B.3 Instructions

The following instructions were given to participants in the videotaped evaluation (see sections 3.1.3 and 5.3).

BioCarta

48. Please open the BioCarta web page (http://www.biocarta.com/genes/index.asp).
49. Find the "arginine biosynthesis" pathway in E. coli. Does the figure allow you to determine which pathways the "arginine biosynthesis" pathway is linked to?
50. Use the above pathway to view information about the enzyme "ornithine carbamoyltransferase".
51. Find the "ATM signaling pathway" in Mus musculus. What does each symbol in the pathway diagram mean (Hint: Look for a legend)?

52. Search for the enzyme "carbamoyl-phosphate synthase" that is responsible for glutamine hydrolysis. Find the names of the pathways this enzyme participates in.

ExPASy

53. Please open the ExPASy web page (http://www.expasy.org/cgi-bin/search-biochem-index).
54. Find the "arginine biosynthesis" pathway in E. coli. Does the figure allow you to determine what the colors and shapes of the arrows in the pathway diagram mean (i.e. is there a legend somewhere)? Does the figure allow you to determine which pathways the "arginine biosynthesis" pathway is linked to?
55. Go to map G7, G8, by clicking on the right arrow from map F8 from question 7 above, and then clicking on the top arrow. Use this pathway (in map G7, G8) to view information about the enzyme "ornithine carbamoyltransferase".
56. Find the insulin receptor. Is there a legend that allows you to determine what each symbol in the pathway diagram mean?
57. Search and find information about the enzyme "carbamoyl-phosphate synthase" that is responsible for glutamine hydrolysis.

KEGG

58. Please open the KEGG web page (http://www.kegg.com/kegg/kegg2.html).
59. Find the "arginine biosynthesis" pathway in E. coli. Does the figure allow you to determine which pathways the "arginine biosynthesis" pathway is linked to?
60. Use the above pathway to view information about the enzyme "ornithine carbamoyltransferase" that catalyzes the ornithine ←→ citrulline reaction.
61. What do enzymes in green boxes represent (Hint: Look for a legend)?
62. Find the MAPK signaling pathway in Homo sapiens.
63. Search for the enzyme "carbamoyl-phosphate synthase" that is responsible for glutamine hydrolysis. Find the names of the pathways this enzyme participates in.
64. Color the "carbamoyl-phosphate synthase" (E.C. # 6.3.4.16) enzyme red and the "ornithine carbamoyltransferase" (E.C. # 2.1.3.3) enzyme pink. View both in the "urea cycle" pathway in E. coli.

WIT

65. Please open the WIT web page (http://wit.mcs.anl.gov/WIT2/).
66. Find the "arginine biosynthesis" pathway in E. coli. Does the figure allow you to determine which pathways the "arginine biosynthesis" pathway is linked to?
67. Use the above pathway to view information about the enzyme "ornithine carbamoyltransferase".
68. Search for the enzyme "carbamoyl-phosphate synthase" that is responsible for glutamine hydrolysis. Find the names of the pathways this enzyme participates in.

MetaCyc

69. Please open MetaCyc.
70. Find all "carbon compounds degradation" pathways in E. coli. How many are there?
71. Find the "arginine biosynthesis II" pathway in E. coli.
72. Find the compound "carbamoyl-phosphate" in the above pathway. Show other pathways that the compound participates in.
73. Find all inhibitors and activators of the enzyme that catalyzes the "carbamoyl-phosphate ←→ $HCO3^{-}$" reaction in the "arginine biosynthesis II" pathway; namely the "carbamoyl phosphate synthase enzyme".
74. As a biochemist, you would like to compare the "arginine biosynthesis II" pathway to "alanine biosynthesis" and "cysteine biosynthesis" pathways. To accomplish this task, you would like to view the above pathways simultaneously. This tool allows you to do so. Try to figure out how.
75. Click on the "Overview Mode". Figure out what each symbol in the overview window means (Hint: Look for a legend).
76. Highlight all amino acid biosynthesis pathways in E. coli.
77. Highlight the pathways that are shared between E. coli and all organisms. How many shared pathways are there?

BioPathVis

1. Please start the BioPathVis application.
2. Find the names of all “Carbohydrate Metabolism” pathways that occur in the “Homo Sapiens” species, “liver” organ, and “cytosol” organelle.
    a. View an overview image of the “Carbohydrate Metabolism” pathways.
    b. View the “Glycolysis / Gluconeogenesis” pathway using the overview.
3. Find the “Nucleotide sugars metabolism” pathway in the “E. coliK-12” species.
    a. What do the colors in the pathway diagram mean (i.e. is there a legend somewhere)?
        i. If there is a legend, hide it.
    b. Does the figure allow you to determine which pathways the “Nucleotide sugars metabolism” pathway is linked to?
        i. If so, then view the pathway “Glycolysis / Gluconeogenesis”.
    c. In the “Nucleotide sugars metabolism” pathway in the “E.coliK-12” species,
        i. Scroll to the upper left corner of the pathway graph. Expand the pathway graph panel horizontally to fit the entire screen.
        ii. View information about the enzyme “dTDPglucose 4,6-dehydratase” that catalyzes the “dTDPglucose <=> dTDP-4-dehydro-6-deoxy-D-glucose + H2O” reaction.
        iii. View information about the reaction “dTDPglucose <=> dTDP-4-dehydro-6-deoxy-D-glucose + H2O”.
        iv. Scroll to the upper left corner of the pathway figure. Zoom out such that you can see the entire pathway.
        v. Hide enzymes and co-substrates from the pathway figure.
4. Click on the “Glycolysis / Gluconeogenesis” tab for the “Homo Sapiens” species, “liver” organ, and “cytosol” organelle.
    a. Save this pathway as a jpeg file, titled “Glycolysis Gluconeogenesis _homo sapiens_liver_cytosol”, under C:\tmp.
    b. View the saved image.

5. Use advanced search to find (simultaneously) the following pathways in the "E. coliK-12" species: "Pyruvate metabolism", "Propanoate metabolism", and "Nucleotide sugars metabolism".
    a. Display the "Nucleotide sugars metabolism" pathway using the search results.
6. Use advanced search to find (simultaneously) all amino acids and lipids (compounds) in the "E.coliK-12" species and "cytosol" organelle.
    a. View information about the compound "L-alanine" using the search results.

B.4 Performance Data

Table B-1: Speed and average of task performance (s) for each user using all tools. N/V (Not Valid) means that the measurement is not valid because considerable help was given to the user or because the user did not accomplish the task correctly. Infinite times are times for tasks that took excessive time (greater than 180 seconds for most tasks) so the user gave up doing them. Other numbers in red italics are not valid because the user only accomplished part of the task

Tool	Task	Time (s) User 1	Time (s) User 2	Time (s) User 3	Time (s) User 4	Time (s) User 5	Avg (s)
Bio-Carta	2.1 Find arginine biosynthesis pathway in E. coli	34	41	26	114	42	51.40
	2.2 Find links on pathway	8	19	81	27	4	27.80
	Total	42	60	107	141	46	79.20
	3. Find enzyme ornithine carbamoyltransferase on figure	1	37	5	3	4	10.00
	4.1 Find ATM signaling pathway in Mus Musculus	60	68	*N/V*	*N/V*	54	61.00
	4.2 Find legend	8	12	*N/V*	*N/V*	53	24.00
	Total	68	80	32	63	107	70.00
	5. Find enzyme carbamoyl phosphate synthetase (glutamine hydrolyzing)	∞	∞	*249*	*N/V*	*182*	∞

Tool	Task	Time (s) User 1	Time (s) User 2	Time (s) User 3	Time (s) User 4	Time (s) User 5	Avg (s)
Ex-PASy	7.1 Find arginine biosynthesis pathway	124	162	116	*238*	93	146.60
	7.2 Find if there is a legend	1	22	35	33	29	24.00
	7.3 Find links on pathway	18	12	6	9	98	28.60
	Total	143	196	157	280	220	199.20
	8. Find enzyme ornithine carbamoyltransferase on figure	125	30	8	26	15	40.80
	9.1 Find insulin receptor	124	91	17	41	32	61.00
	9.2 Find if there is a legend	28	34	74	19	5	32.00
	Total	152	125	91	60	37	93.00
	10. Find enzyme carbamoyl phosphate synthase (glutamine hydrolyzing)	30	68	59	129	49	67.00
Meta-Cyc	12. Find all carbon compounds degradation pathways in E. coli and how many there are	134	65	237	115	66	123.40
	13. Find arginine biosynthesis II pathway in E. coli	41	54	29	58	15	39.40
	14.1 Find compound carbamoyl phosphate in pathway figure	23	13	3	15	8	12.40
	14.2 Find other pathways compound participates in	5	16	37	18	15	18.20
	Total	28	29	40	33	23	30.60
	15. Find all activators and inhibitors of carbamoyl phosphate synthase	38	74	32	150	60	70.80
	16. View pathways simultaneously (arginine, alanine, & cysteine biosynthesis)	310	161	*327*	*220*	78	219.20
	17. Find a legend for the symbols in the overview mode.	142	67	82	66	43	80.00
	18. Highlight all amino acid biosynthesis pathways in E. coli	64	354	131	99	152	160.00

Tool	Task	Time (s) User 1	Time (s) User 2	Time (s) User 3	Time (s) User 4	Time (s) User 5	Avg (s)
Meta-Cyc	19. Highlight pathways shared between E. coli and all organisms. How many?	*56*	136	61	*35*	50	67.60
KEGG	21.1 Find arginine biosynthesis pathway in E. coli	132	281	90	*135*	133	159.00
	21.2 Find links on pathway	15	70	13	*15*	11	27.00
	Total	147	351	103	*150*	144	179.00
	22. Find enzyme ornithine carbamoyltransferase on figure	23	45	10	65	15	31.60
	23. Find a legend to specify what enzymes in green boxes represent	21	32	33	29	26	28.20
	24. Find MAPK signaling pathway in Homo Sapiens	38	57	31	27	20	34.60
	25.1 Find enzyme carbamoyl phosphate synthase (glutamine hydrolyzing)	73	194	60	261	92	136.00
	25.2 Find pathways this enzyme participates in	17	34	28	22	18	23.80
	Total	90	228	88	283	110	159.80
	26. Color 6.3.4.16 enzyme red and 2.1.3.3 enzyme pink. View both in urea cycle.	279	331	*261*	∞	∞	∞
WIT	28.1 Find arginine biosynthesis pathway in E. coli	194	*284*	*103*	*173*	160	182.80
	28.2 Find links on pathway	18	35	67	38	10	33.60
	Total	212	*319*	*170*	*211*	170	216.40
	29. Find enzyme ornithine carbamoyltransferase on figure	15	19	12	88	17	30.20
	30.1 Find enzyme carbamoyl phosphate synthase (glutamine hydrolyzing)	50	118	73	252	149	128.40
	30.2 Find pathways this enzyme participates in	30	78	*N/A*	78	47	58.00
	Total	80	196	*73*	330	196	175.00

Tool	Task	Time (s) User 1	Time (s) User 2	Time (s) User 3	Time (s) User 4	Time (s) User 5	Avg (s)
Bio-Path-Vis	2. Find names of all "Carbohydrate Metabolism" pathways that occur in "Homo Sapiens", "liver", "cytosol".	26	17	20	60	91	42.80
	2.1 View overview image of "Carbohydrate Metabolism" pathways.	5	1	1	9	104	24.00
	2.2 View "Glycolysis / Gluconeogenesis" pathway using overview.	6	12	10	18	14	12.00
	3. Find "Nucleotide sugars metabolism" pathway in "E. coliK-12".	38	18	9	32	21	23.60
	3.1 Find if there is a legend.	2	1	8	5	8	4.80
	3.1.1 Hide legend.	36	8	18	27	4	18.60
	3.2 Find linked pathways.	21	1	4	26	99	30.2
	3.2. View linked pathway "Glycolysis / Gluconeogenesis".	11	3	9	1	9	6.60
	3.3 In "Nucleotide sugars metabolism" pathway in "E.coliK-12",						
	3.3.1 Expand pathway graph panel horizontally to fit entire screen.	4	52	1	44	37	27.60
	3.3.2 View information about enzyme "dTDPglucose 4,6-dehydratase".	109	18	21	18	66	46.40
	3.3.3 View information about reaction "dTDPglucose <=> dTDP-4-dehydro-6-deoxy-D-glucose + H2O".	12	136	30	41	8	45.40
	3.3.4 Zoom out to see entire pathway.	17	3	1	9	8	7.60
	3.3.5 Hide enzymes and co-substrates.	3	2	3	4	21	6.60
	4. Click on "Glycolysis / Gluconeogenesis" tab for "Homo Sapiens", "liver", "cytosol".	22	4	4	28	4	12.40
	4.1 Save this pathway as a jpeg file.	24	27	11	45	18	25.00
	4.2 View saved image.	10	41	11	37	30	25.80

Tool	Task	Time (s) User 1	Time (s) User 2	Time (s) User 3	Time (s) User 4	Time (s) User 5	Avg (s)
Bio-Path-Vis	5. Use advanced search to find in "E. coliK-12": "Pyruvate metabolism", "Propanoate metabolism", and "Nucleotide sugars metabolism".	168	78	68	61	93	93.60
	5.1 Display "Nucleotide sugars metabolism" using search results.	1	1	1	10	7	4.00
	6. Use advanced search to find all amino acids and lipids (compounds) in "E.coliK-12", "cytosol".	43	52	17	61	140	62.60
	6.1 View information about "L-alanine" using search results.	4	9	1	13	5	6.40

Table B-2: T-test comparing average of task performance for each tool pair ($\alpha = 0.05$)

Task	Tool	$\bar{x}$	s^2	n	df	t_{stat}	$t_{0.05}$ [36]	H_a	Test	Conclusion
1	BioCarta	51.40	1265.80	5	7	-3.377	-1.895	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	ExPASy	123.75	822.92	4						
	BioCarta	51.40	1265.80	5	6	0.675	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	MetaCyc	39.40	316.30	5						
	BioCarta	51.40	1265.80	5	4	-2.402	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	KEGG	159.00	7016.67	4						
	BioCarta	51.40	1265.80	5	3	-5.394	-2.353	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	WIT	177.00	578.00	2						
	BioCarta	51.40	1265.80	5	5	1.662	2.015	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	23.60	123.30	5						
	ExPASy	123.75	822.92	4	5	5.143	2.015	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	MetaCyc	39.40	316.30	5						
	ExPASy	123.75	822.92	4	4	-0.796	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	KEGG	159.00	7016.67	4						
	ExPASy	123.75	822.92	4	3	-2.394	-2.353	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	WIT	177.00	578.00	2						

Task	Tool	$\bar{x}$	s^2	n	df	t_{stat}	$t_{0.05}$ [36]	H_a	Test	Conclusion
1	ExPASy	123.75	822.92	4	4	6.572	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	23.60	123.30	5						
	MetaCyc	39.40	316.30	5	3	-2.805	-2.353	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	KEGG	159.00	7016.67	4						
	MetaCyc	39.40	316.30	5	1	-7.331	-6.314	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	WIT	177.00	578.00	2						
	MetaCyc	39.40	316.30	5	7	1.668	1.894	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	23.60	123.30	5						
	KEGG	159.00	7016.67	4	4	-0.398	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	177.00	578.00	2						
	KEGG	159.00	7016.67	4	3	3.209	2.353	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	23.60	123.30	5						
	WIT	177.00	578.00	2	1	8.637	6.314	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	23.60	123.30	5						
2	BioCarta	27.80	966.70	5	8	-0.036	-1.860	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	ExPASy	28.60	1524.80	5						
	BioCarta	27.80	966.70	5	7	0.028	1.895	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	KEGG	27.25	814.92	4						
	BioCarta	27.80	966.70	5	7	-0.340	-1.895	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	33.60	484.30	5						
	BioCarta	27.80	966.70	5	8	-0.106	-1.856	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	30.20	1593.70	5						
	ExPASy	28.60	1524.80	5	7	0.060	1.895	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	KEGG	27.25	814.92	4						
	ExPASy	28.60	1524.80	5	6	-0.249	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	33.60	484.30	5						
	ExPASy	28.60	1524.80	5	8	-0.064	-1.860	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	30.20	1593.70	5						
	KEGG	27.25	814.92	4	6	-0.366	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	33.60	484.30	5						
	KEGG	27.25	814.92	4	7	-0.129	-1.894	$\mu_1 - \mu_2 < 0$	Reject H_o if	Accept H_o

Task	Tool	$\bar{x}$	s^2	n	df	t_{stat}	$t_{0.05}$ [36]	H_a	Test	Conclusion
2	BioPathVis	30.20	1593.70	5					$t_{stat} < t_{0.05}$	
	WIT	33.60	484.30	5	6	0.167	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	30.20	1593.70	5						
3	BioCarta	10.00	230.00	5	5	-1.371	-2.015	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	ExPASy	40.80	2291.70	5						
	BioCarta	10.00	230.00	5	6	-0.317	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	MetaCyc	12.40	56.80	5						
	BioCarta	10.00	230.00	5	7	-1.755	-1.895	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	KEGG	31.60	527.80	5						
	BioCarta	10.00	230.00	5	6	-1.262	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	30.20	1050.70	5						
	BioCarta	10.00	230.00	5	5	-1.882	-2.015	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	46.40	1640.30	5						
	ExPASy	40.80	2291.70	5	4	1.310	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	MetaCyc	12.40	56.80	5						
	ExPASy	40.80	2291.70	5	6	0.387	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	KEGG	31.60	527.80	5						
	ExPASy	40.80	2291.70	5	7	0.410	1.895	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	WIT	30.20	1050.70	5						
	ExPASy	40.80	2291.70	5	8	-0.200	-1.860	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	46.40	1640.30	5						
	MetaCyc	12.40	56.80	5	5	-1.776	-2.015	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	KEGG	31.60	527.80	5						
	MetaCyc	12.40	56.80	5	4	-1.196	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	30.20	1050.70	5						
	MetaCyc	12.40	56.80	5	4	-1.845	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	46.40	1640.30	5						
	KEGG	31.60	527.80	5	7	0.079	1.895	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	WIT	30.20	1050.70	5						
	KEGG	31.60	527.80	5	6	-0.711	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	46.40	1640.30	5						

Task	Tool	$\bar{x}$	s^2	n	df	t_{stat}	$t_{0.05}$ [36]	H_a	Test	Conclusion
3	WIT	30.20	1050.70	5	8	-0.698	-1.860	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	BioPathVis	46.40	1640.30	5						
4	BioCarta	60.67	49.33	3	4	-0.016	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	ExPASy	61.00	2011.50	5						
	BioCarta	60.67	49.33	3	6	3.474	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	KEGG	34.60	199.30	5						
	ExPASy	61.00	2011.50	5	5	1.255	2.015	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	KEGG	34.60	199.30	5						
5	BioCarta	24.33	620.33	3	4	-0.415	-2.132	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	ExPASy	32.00	670.50	5						
	BioCarta	24.33	620.33	3	6	-2.525	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	MetaCyc	80.00	1395.50	5						
	BioCarta	24.33	620.33	3	2	-0.266	-2.920	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	KEGG	28.20	23.70	5						
	BioCarta	24.33	620.33	3	2	1.351	2.920	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	4.80	10.70	5						
	ExPASy	32.00	670.50	5	7	-2.361	-1.895	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	MetaCyc	80.00	1395.50	5						
	ExPASy	32.00	670.50	5	4	0.322	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	KEGG	28.20	23.70	5						
	ExPASy	32.00	670.50	5	4	2.330	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	4.80	10.70	5						
	MetaCyc	80.00	1395.50	5	4	3.075	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	KEGG	28.20	23.70	5						
	MetaCyc	80.00	1395.50	5	4	4.484	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	4.80	10.70	5						
	KEGG	28.20	23.70	5	7	8.921	1.894	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	4.80	10.70	5						
6	ExPASy	67.00	1400.50	5	5	-1.620	-2.015	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	KEGG	136.00	7667.50	5						
	ExPASy	67.00	1400.50	5	6	-1.569	-1.943	$\mu_1 - \mu_2 < 0$	Reject H_o if	Accept H_o

Task	Tool	$\bar{x}$	s^2	n	df	t_{stat}	$t_{0.05}$ [36]	H_a	Test	Conclusion
6	WIT	128.40	6256.30	5					$t_{stat} < t_{0.05}$	
	ExPASy	67.00	1400.50	5	8	0.165	1.860	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	62.60	2142.30	5						
	KEGG	136.00	7667.50	5	8	0.144	1.860	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	WIT	128.40	6256.30	5						
	KEGG	136.00	7667.50	5	6	1.657	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	62.60	2142.30	5						
	WIT	128.40	6256.30	5	6	1.605	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	62.60	2142.30	5						
7	MetaCyc	44.50	559.63	5	5	1.873	2.015	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	KEGG	23.80	51.20	5						
	MetaCyc	44.50	559.63	5	7	-0.863	-1.895	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	Accept H_o
	WIT	58.25	568.25	4						
	MetaCyc	44.50	559.63	5	4	3.533	2.132	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	6.40	21.80	5						
	KEGG	23.80	51.20	5	3	-2.791	-2.353	$\mu_1 - \mu_2 < 0$	Reject H_o if $t_{stat} < t_{0.05}$	***Reject H_o***
	WIT	58.25	568.25	4						
	KEGG	23.80	51.20	5	7	4.554	1.894	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	6.40	21.80	5						
	WIT	58.25	568.25	4	3	4.285	2.353	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	6.40	21.80	5						
8	MetaCyc	123.40	4948.30	5	6	2.332	1.943	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	***Reject H_o***
	BioPathVis	42.80	1021.70	5						
9	MetaCyc	183.00	13819.00	3	2	1.267	2.920	$\mu_1 - \mu_2 > 0$	Reject H_o if $t_{stat} > t_{0.05}$	Accept H_o
	BioPathVis	93.60	1874.30	5						

Table B-3: ANOVA test testing equality of task performance means for existing tools (α = 0.05)

Task	Tool	$\bar{x}$	s^2	n	SST	df SST	SSE	df SSE	MST	MSE	F	$F_{0.05}$	Conclusion
1	BioCarta	51.40	1265.80	5	58023.80	4	30425.15	15	14505.95	2028.34	7.15	3.05	***Reject H_o***
	ExPASy	123.75	822.92	4									
	MetaCyc	39.40	316.30	5									
	KEGG	159.00	7016.67	4									
	WIT	177.00	578.00	2									
2	BioCarta	27.80	966.70	5	122.68	3	14347.95	15	40.89	956.53	0.043	3.29	Accept H_o
	ExPASy	28.60	1524.80	5									
	KEGG	27.25	814.92	4									
	WIT	33.60	484.30	5									
3	BioCarta	10.00	230.00	5	3520.00	4	16628.00	20	880.00	831.40	1.06	2.87	Accept H_o
	ExPASy	40.80	2291.70	5									
	MetaCyc	12.40	56.80	5									
	KEGG	31.60	527.80	5									
	WIT	30.20	1050.70	5									
4	BioCarta	60.67	49.33	3	2124.44	2	8941.87	10	1062.22	894.19	1.19	4.10	Accept H_o
	ExPASy	61.00	2011.50	5									
	KEGG	34.60	199.30	5									
5	BioCarta	24.33	620.33	3	9590.53	3	9599.47	14	3196.84	685.68	4.66	3.34	***Reject H_o***
	ExPASy	32.00	670.50	5									
	MetaCyc	80.00	1395.50	5									
	KEGG	28.20	23.70	5									
6	ExPASy	67.00	1400.50	5	14314.53	2	61297.20	12	7157.27	5108.10	1.40	3.88	Accept H_o
	KEGG	136.00	7667.50	5									
	WIT	128.40	6256.30	5									
7	MetaCyc	44.50	559.62	5	2730.68	2	4148.05	11	1365.34	377.10	3.62	3.98	Accept H_o
	KEGG	23.80	51.20	5									
	WIT	58.25	568.25	4									

Table B-4: ANOVA test testing equality of task performance means for all tools (α = 0.05)

Task	Tool	$\bar{x}$	s^2	n	SST	df SST	SSE	df SSE	MST	MSE	F	$F_{0.05}$	Conclusion
1	BioCarta	51.40	1265.80	5	79544.69	5	30954.35	9	15908.94	1629.18	9.76	2.74	***Reject H_o***
	ExPASy	123.75	822.92	4									
	MetaCyc	39.40	316.30	5									
	KEGG	159.00	7016.67	4									
	WIT	177.00	578.00	2									
	BioPathVis	23.60	132.30	5									
2	BioCarta	27.80	966.70	5	125.08	4	20722.75	19	31.27	1090.67	0.029	2.90	Accept H_o
	ExPASy	28.60	1524.80	5									
	KEGG	27.25	814.92	4									
	WIT	33.60	484.30	5									
	BioPathVis	30.20	1593.70	5									
3	BioCarta	10.00	230.00	5	5428.17	5	23189.20	24	1085.63	966.22	1.12	2.62	Accept H_o
	ExPASy	40.80	2291.70	5									
	MetaCyc	12.40	56.80	5									
	KEGG	31.60	527.80	5									
	WIT	30.20	1050.70	5									
	BioPathVis	46.40	1640.30	5									
4	BioCarta	60.67	49.33	3	2124.44	2	8941.87	10	1062.22	894.19	1.19	4.10	Accept H_o
	ExPASy	61.00	2011.50	5									
	KEGG	34.60	199.30	5									
5	BioCarta	24.33	620.33	3	15300.60	4	9642.27	18	3825.15	535.682	7.14	2.93	***Reject H_o***
	ExPASy	32.00	670.50	5									
	MetaCyc	80.00	1395.50	5									
	KEGG	28.20	23.70	5									
	BioPathVis	4.80	10.70	5									
6	ExPASy	67.00	1400.50	5	22906.60	3	69866.40	16	7635.53	4366.65	1.74	3.24	Accept H_o
	KEGG	136.00	7667.50	5									
	WIT	128.40	6256.30	5									

Task	Tool	$\bar{x}$	s^2	n	SST	df SST	SSE	df SSE	MST	MSE	F	$F_{0.05}$	Conclusion
6	BioPathVis	62.60	2142.30	5									
7	MetaCyc	44.50	559.62	5	7150.38	3	4235.25	15	2383.46	282.35	8.44	3.29	***Reject H_o***
	KEGG	23.80	51.20	5									
	WIT	58.25	568.25	4									
	BioPathVis	6.40	21.80	5									
8	MetaCyc	123.40	4948.30	5	16240.90	1	23880.00	8	16240.90	2985.00	5.44	5.32	***Reject H_o***
	BioPathVis	42.80	1021.70	5									
9	MetaCyc	183.00	13819.00	3	14985.68	1	35135.20	6	14985.68	5855.87	2.56	5.99	Accept H_o
	BioPathVis	93.60	1874.30	5									

B.5 Malfunction Data

Table B-5: A list of videotaped evaluation malfunctions (VM) in BioCarta, including task # that uncovered malfunction (See Appendix B.3), user #, malfunction description, UI guidelines violated, severity level from 1 to 5 (See Table 3-1), and recommendations for change

VM #	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Feedback and Error Handling					
B1	2, 4/ User 1, 3, 5 4/ All users	The user cannot tell from the pathway figure which organism or species the pathway belongs to. Besides the pathway figure, there are two buttons; one is titled "Humans" and the other "Other Species". When clicking on "Other Species", the user is not informed which species the figure belongs to.	-Uniquely identify each state (Make each change of state immediately clear)	4	-Provide a combo box on top of the figure that states "species" besides it. Allow the user to choose the appropriate species from the combo box. -Confirm the species change in the title of the figure.

VM #	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Font					
B2	4/ User 1, 5	It was not very easy to find a legend. The legend link is located at the top, along with other links, and in small font. User 5 thought that the legend is the text under the figure, which is just a description of the pathway, but not a legend.	-Make each state clear to the user.	5	-Provide a legend under figure or make the legend link more visible (perhaps right on the figure) and in bigger font.
Malfunctions Related to Information Query Interfaces					
B3	2/ User 1, 4, 5	Searching for a pathway by typing in the pathway name requires that the user types in the keywords in the exact order they appear in the pathway name. For example "arginine biosynthesis" does not work, whereas "Biosynthesis of arginine" works. The search should be more flexible.	-The system should allow the user to search for any items containing given keywords.	4	-Make the search more flexible by allowing the system to search using any combination of keywords.
B4	2/ User 2	User considered the "New Pathway" link first to search for arginine biosynthesis. "New Pathway" should not be placed in the section that allows the user to browse pathways, because it does not belong there and this could mislead the user.	-Ensure all items correspond to the name of the menu.	4	-Do not place "New Pathway" under "BROWSE PATHWAYS BY CATEGORY", since all the items under this category should allow the user to search. Create a new category for pathway editing.

VM #	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
B5	2, 4/ User 1, 3 4/ User 2 2/ User 4	When the user wants to browse pathways, the pathways are listed in alphabetical order. The list is huge and searching for a name is not intuitive. For instance, if one wants to look for arginine metabolism, it is listed under "Biosynthesis of arginine". It is not intuitive that the user would search under the letter B (for Biosynthesis), as opposed to A (for arginine).	-Order items and groups logically (Use natural ordering where possible and group alphabetically as a last resort)	4	-Categorize pathways (e.g. carbohydrate metabolism, amino acid metabolism, etc.) -List pathways under each category in alphabetical order. -Make the pathway name more descriptive of the main compound that is synthesized or degraded. (e.g. "Arginine Biosynthesis", as opposed to "Biosynthesis of Arginine").
B6	3/ All users	To view info about an enzyme, the user needs to click on the enzyme, then on the different links on the enzyme info page. User 1 thought "I can't really find information on it. Well, I'd have to go into different sites". User 1 is concerned with "having to jump around 20 websites". It's not clear what kind of info each link displays. User 3 commented that "it's just like another search". User 4 thought that the search results don't give you much information.	-Allow the user to find the information easily and quickly. -Do not confuse the user. -Reduce the user's cognitive efforts.	5	-Provide a brief description of the most important information about the enzyme in the main page. -Provide links to additional information that experts may need. The links should have a description of the type of information they display.

VM #	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
B7	5/ User 1, 2	It is not intuitive that the "Genes" tab and "Gene Name" allow one to search for enzymes. These titles are misleading. The user considered typing the enzyme name in the "Pathway Name" field, but not in "Gene Name" field. Even after giving hint about looking for a compound, user 1 could not find it and eventually gave up. User 2 said "I don't think I want a gene. I want an enzyme". Even after trying to browse through pathways and looking for something that contains the enzyme, the user came back to home page and said "I don't want gene name". The user said "I'm lost" and expressed frustration.	-Keep labels unambiguous.	3	-Change "Genes" tab to "Compound" tab. -Change "Gene Name" to "Compound Name". -Provide a combo box, with a name "Compound Type" and allow the user to choose whether to search for an enzyme, substrate, etc.
B8	5/ User 1, 3, 5	When searching for an enzyme using the "Gene Name" field, the search results display a link that states "reactions that feed amino groups in the urea cycle". We expect to find enzyme names in the search results and not reactions. Apparently, the "Gene Name" field allows searching for pathways by the gene they contain. A more descriptive name should be used.	-Keep labels unambiguous.	4	-Change "Gene Name" to "Search for Pathway Containing Compound". -Change "Pathway Name" to "Search for Pathway" to be consistent with the other field.

VM #	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
B9	5/ User 1, 3	Typing in the enzyme using the "Pathway" tab in the "Gene Name" field displays different results from searching for the enzyme using the "Genes" tab fields. Typing in the enzyme in the "Gene Name" displays a link that states "reactions that feed amino groups in the urea cycle". When clicking on this reaction, a pathway figure is displayed with the enzyme. When clicking on the enzyme, the information page displayed is different from the one displayed when searching for the same enzyme using the "Gene" tab.	-Be consistent	4	-Provide consistent search results.
B10	4, 5/ User 1, 2, 5	To go back to pathway search page, the user clicked on "back" button several times.	-Provide a quick way to go back to main search.	5	-Provide a link to go to the main pathway page quickly, instead of having to press on the back button several times.

Table B-6: A list of videotaped evaluation malfunctions (VM) in ExPASy, including task # that uncovered the malfunction (See Appendix B.3), user #, malfunction description, UI guidelines violated, severity level from 1 to 5 (See Table 3-1), and recommendations for change

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Feedback and Error Handling					
E1	7/ User 3	The error message generated when typing in the wrong pathway name is ambiguous. It states that the pathway is not available in the current version, which made the user wonder if it's available in a different version.	-Provide details about errors. -State error messages in terms of conceptual model.	4	-Change the error message to describe what the error is about. One possibility is "No pathway is found that contains the compound you have typed. Please enter another compound name".
Malfunctions Related to Information Query Interfaces					
E2	7/ User 1, 2, 4, 5 9, 10/ All users	This tool does not allow one to search for pathways by name. When the user typed "arginine biosynthesis", an error page is displayed. Keyword search only allows searching for entries (e.g. compounds) in the pathway map.	-Provide flexibility.	4	-Make the search more flexible. -Allow users to search for pathways by name. -Allow users to refine search (e.g. define searching category, such as pathways, compounds, etc.).
E3	7/ User 1, 2, 4	The quick search combo box located at the top of the main page does not allow the user to search for pathways or entries in the pathways. User 1 and 2 tried examining combo box and gave up. User 4 tried searching for "arginine biosynthesis" using "ExPASy web site" option, but the search results returned the occurrences of the words in text, not pathway figures.	-Provide flexibility.	4	-The quick search in the pathway tool should allow the user to search for pathways, since the main purpose of the pathway tool is displaying pathways.

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
E4	7/ User 1, 2, 5	It is not clear that the "map" link shows figures. The search results of "arginine" returned a list of compounds, with a "map" link and an "enzyme" link underneath each. User 1 was not sure if the user was on right track. User 2 and 5 thought that the results only gave a list of enzymes and went back to the search page. One user said "here it says available map, but I'm not sure what it does".	-Keep labels unambiguous.	3	-Change the "map" link to "Show pathway figure". -If the tool is changed to allow users to refine the search by selecting whether the user wants to search for pathways or compounds, then the search results of pathways should only display links to pathway figures.
E5	7/ User 1, 2	This tool does not allow the user to find pathways in different organisms. The tool shows different colors of arrows that mean different organism classes, but there is no legend that states so.	-Provide necessary functionality. -If help is necessary, provide it.	3	-Allow the user to display pathways in different species or organisms. -Provide a legend to show that the colors of the arrows mean different organism classes.
E6	7, 8/ All users 9/ User 2, 3, 4	No legend to state what the colors mean. A legend is necessary because there are different types and colors of arrows and compounds and it would be easier for the user to search for information when the user knows what the colors mean.	-Do not confuse the user. If help is necessary, provide it.	3	-Provide a clear legend of what each color and shape in the figure means.

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
E7	7/ User 2	When clicking on outside arrows, it only allows you to view 2 maps at a time, sometimes vertically and sometimes horizontally. Although it's good to be able to view 2 maps at a time, user 2 commented that the user had lost what the user had before when clicking on arrows.	-Allow the user to get a clear picture of what the user is looking for.	5	-Allow the user to display more than one map at a time (perhaps 4), because this will make it easier for the user to see all the pathways surrounding a certain compound.
E8	7/ User 3, 4	The overview of metabolic pathways picture is too huge and the information it contains is too huge for the user to be able to search for information by just looking at the figure. The picture is useless and the user gave up looking for information using it.	-Do not confuse the user. -The system must allow the user to accomplish tasks without much cognitive effort.	3	-Allow the user to zoom in and out the overview picture. -Allow the user to navigate through the overview picture. -Allow the user to search for information on the overview figure.
E9	9/ All users	No easy way to go back to the main search page. There is a link at the top saying main page, but it doesn't take you directly to pathway search.	-Provide a quick way to go back to main search.	5	-Provide a link to go to the main page quickly, instead of having to press on the back button several times.
E10	7, 9/ All users	Map names are ambiguous. The user cannot determine what the map name (such as U4) means by just considering the name.	-Keep labels unambiguous.	4	-Provide a more meaningful map name, or provide a description of the map name.
E11	10/ User 1, 5	The user did not notice that the "Enzyme" link in the search results allows one to search for enzymes. The user went to the map and then looked for the enzyme, as opposed to using the link.	-Allow the user to keep the correct mental model.	4	-If this tool allows the user to choose whether the user is searching for enzymes or pathways, then this problem will not occur.

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
E12	10/ User 1	BUG (Incorrect info): If you click on map that contains the carbamoyl-phosphate synthase that is responsible for glutamine hydrolysis, then on carbamoyl-phosphate synthase on the figure, it takes you to the information page for the one for ammonia, and not for glutamine hydrolysis.	-Provide correct and consistent information.	4	-Provide the same information when the user searches for the same information in different ways.
E13	10/ User 3	The user thought that in order to find an enzyme, the user has to click on the enzyme database. Despite the fact that the user clicked on the "ENZYME" link to the database, the user did not actually use the page to search. The user used the top search combo box to do the search.	-Allow the user to keep the correct mental model and identify the state the user is at. -Provide feedback and help on items.	5	-Allow the user to refine the search and choose what the user is searching for in the main page. The main page can have one link for pathways and another for compounds. The quick search box should be at the very top.

Table B-7: A list of videotaped evaluation malfunctions (VM) in KEGG, including task # that uncovered the malfunction (See Appendix B.3), user #, malfunction description, UI guidelines violated, severity level from 1 to 5 (See Table 3-1), and recommendations for change

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Feedback and Error handling					
K1	21/ User 1, 2, 4	The figure does not emphasize which organism it belongs to. The figure shown at the beginning of the search is for the general pathway and the user later must choose the organism from the combo box at the top and hit "Exec". The user thought that choosing an organism would be done before viewing a pathway. User 2 thought the tool does not provide searching for organisms and eventually gave up.	-Provide necessary functionality. -Allow the user to know which state the user is at, without much cognitive effort.	4	-State the species in the title. -Confirm the change of species in the title. -Allow the user to choose the species before displaying the pathway, perhaps in a combo box.
K2	26/ User 1	Ambiguous error messages. When the user tried entering the two enzyme E.C. numbers separated by one line, changing the second field to "red pink", and hitting "Exec", the error message stated "Invalid organism name". Also, when the user tried "6.3.4.16, 2.1.3.3" and "red,pink", the error message stated that the "Following EC/Compound/Gene(s) was/were not found".	-Provide as much detail as you can about the errors (cause(s) and solution(s)). -Display error messages as soon as possible, even before the user finishes entering data on a form.	3	-Display error message right on the search page, in red. - The first error message should state "The red pink field is invalid. Enter color(s) separated by commas". The second error message should've stated "Invalid Entry. Enter the enzyme name(s), followed by space(s), followed by the color. Enzyme names should be separated by one blank line".
Malfunctions Related to Information Query Interfaces					

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
K3	21/ User 2	Typing “arginine” in the quick search combo box displays enzymes only. The search could have been more flexible to allow user to search for pathways as well.	-Provide searching flexibility.	4	-Make the search more flexible. -Allow users to refine search (e.g. define searching category, such as pathways, compounds, etc.).
K4	21/ User 1, 2, 4	The pathway figures are very crowded, making it hard to find information. User 1 commented that it’s hard to find information, or links, since the other linked pathways are not highlighted. User 2 had a hard time following the arrows to see if the pathway is linked to others. User 4 could not even see the adjacent pathway links.	-Present information in a way that reduces cognitive efforts.	4	-Provide a legend to explain what each shape and color in the figure means. -Present different shapes and colors to represent different things. However, be careful not to use too many colors and to be consistent with the use of colors (for more information, see [21]).
K5	21/ User 1, 3, 4, 5	Searching for pathways by category lists the pathways and the user has to search through. It would be easier if the user can type a partial pathway name, which takes the user to the matches. User 5 didn’t want to look through the list and went back to look for another method to search for a pathway by name.	-Provide searching flexibility.	5	-Make the search more flexible. -Allow users to search for a pathway by typing a partial name.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
K6	21/ User 5 24/ User 4 25/ User 1, 5, 4	The user experienced difficulties doing simple searches because the links that allow the user to search have ambiguous names or are not easy to find. In order to search for pathways, the user must click on the "PATHWAY" link or the "Metabolic Pathway" or "Regulatory Pathway" links. In order to search for enzymes, the user must click on the "ENZYME" link. Doing searches using the above links is confusing and not intuitive. The user did not know that these links may take to search pages. The user often clicked on the wrong link to search.	-Labels must be unambiguous.	3	-Make the labels more meaningful. Instead of "PATHWAY" link, use "Search for pathway". Instead of the "ENZYME" link, use "Search for Enzyme". -Provide a quick search capability at the top that allows the user to search for anything (pathways, enzymes, etc). -Reorganize the main page so that the most important and repetitive tasks (such as searching for enzymes and pathways) are placed at the top.
K7	21/ User 5	Looking for the arginine biosynthesis pathway using DBGet "PATHWAY" link displayed lots of results in different organism. It would be better if the user can choose to refine the search earlier so the user does not have to go through the list.	-Provide searching flexibility.	4	-Make the search more flexible. -Allow users to select the organism in the quick search.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
K8	22/ User 1, 2, 4, 5	The user cannot search for enzymes on a figure, unless the user knows the E.C. number of the enzyme or the reaction the enzyme catalyzes. Thus, given an enzyme name, the user was not capable of determining where it is on the figure.	-Avoid ambiguity.	3	-Provide the enzyme name and the E.C. number. If a choice has to be made between the two, then provide the name in the figure and provide the E.C number when the user clicks on the enzyme for more info.
K9	22/ User 2	The enzyme information page is not nicely presented and does not have as much information according to the user.	-Present information in a way that reduces cognitive efforts.	5	-Present the most important information first, with clear distinct sections.
K10	23/ All users	No legend is present on the pathway diagram to determine what the colors in the figure mean. The user must go back to the main page and search there, which is not intuitive and time consuming. None of the users guessed that there is a legend (which is found in the "Pathway map" link in the main page).	-Present information in a way that reduces cognitive efforts. -Provide help.	4	-Provide a legend right on the figure itself.
K11	24/ User 2	The user must hit "Exec" after changing an organism in the combo box. User 2 assumed the page will change automatically upon changing the organism.	-Avoid redundancy.	5	-Change the figure to the chosen species right when the user changes the combo box. -Remove the "Exec" button.
K12	25/ User 1	Searching enzymes using DBGet is confusing. The user found the "bfind mode" and "bget mode" radio buttons confusing. No help was provided.	-Keep labels unambiguous.	4	-Provide help (tool tips) on the radio buttons.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
K13	25/ User 1	User commented that the user does not know what DBGet means. The label is ambiguous.	-Keep labels unambiguous.	4	-Change "DBGet" to "Search for enzymes" or whatever searches the user is performing.
K14	25/ User 2, 4	The "ENZYME" link that allows searching for enzymes is found under "Hierarchical Classification", which is not clear to user.	-Keep labels unambiguous. -Present information in a way that reduces cognitive efforts.	3	-Change the "Hierarchical Classification" title to something more meaningful. -Change the "ENZYME" link to "Search for enzyme".
K15	26/ User 1, 2, 4, 5	The link for coloring objects in pathways should not be placed with links for searching. Some users gave up this task. Most users attempted to look for the pathway first and see if there is an option to color the enzyme later.	-Make sure that the items in a group belong to that group. -Order items and groups logically.	3	-Remove the coloring link from the search area. Have a separate section for coloring. -Allow users to do colorings on the pathway figure itself, instead of placing the links on the main page.
K16	26/ User 1, 2	Coloring page fields are not intuitive. Some labels are ambiguous and some require users to enter information in a specific format. Help is provided in a separate link.	-Provide an example of how to enter info right on the field. -Provide tool tips for each field.	4	-Provide help (tool tips) on each field in the form. -When a specific format is required, provide an example of how to enter information next to the field.
K17	26/ User 3, 4	Clicking on "Color objects in Pathway Map" displays a page titled "Color Genes in the Pathway map", which is confusing to user. The user thought that the user is not on the right track and went back to look for something else. When asked why, User 3 said that the user wants to color enzymes not genes.	-Be consistent.	3	-Change the "Color Genes in the Pathway map" title and link to "Color compounds in the pathway figure" to allow consistency.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
K18	26/ User 3	User 3 found that the "Color for the reference in map" field in the "Color Genes in the Pathway Map" page confusing.	-Keep labels unambiguous.	4	-Change the title to "Color of compound in pathway figure". -Provide help (tool tips) on the field when the cursor is placed on top of it.
K19	26/ User 1, 2	User 3 found "Default color for gene(s) map" field confusing.	-Keep labels unambiguous.	4	-Change the title to "Default color for compound(s) in pathway figure". -Provide help (tool tips) on the field when the cursor is placed on top of it.
Malfunctions Related to On-Line Help					
K20	26/ User 1, 2, 3	The help link on coloring objects is found at the bottom of the coloring objects page. The link is not clear to user. The user spent a considerable amount of time on the help window to understand how to enter info.	-Make help easy to access. -Organize help around tasks and goals.	4	-Provide help (tool tips) on each field in the form. -Make the information in the help page concise and complete.
K21	26/ User 1	The help window tells the user to enter a tab between the enzyme number and the color. However, the tab button takes you to the next field. The user commented that tab does not work.	-Make help accurate. -Do not mislead the user.	5	-Remove the fact that the user can enter a tab between the enzyme names. It misleads the user into thinking that clicking on the tab button inserts a tab.
K22	26/ User 1	It's annoying to switch pages back and forth between help window page and search window page. Perhaps having a popup or putting the information right beside the input window is better.		5	-Provide a popup when clicking on the help link or provide help next to the fields.

Table B-8: A list of videotaped evaluation malfunctions (VM) in MetaCyc, including task # that uncovered the malfunction (See Appendix B.3), user #, malfunction description, UI guidelines violated, severity level from 1 to 5 (See Table 3-1), and recommendations for change

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Color					
M1	15/ User 2, 3	The user experienced difficulties navigating through description pages of compounds, genes, enzymes, substrates, and products. These pages are full of information with too many bright colors. User 2 commented that "I don't like the color. It's not easy on the eyes". The user did not like the combination of orange on white. User 3 did not like the use of small font.	-Use color sparingly (at most 2-4 colors and only color a few items). -Provide a legend to indicate what colors mean.	4	-Use at most 2-4 colors. -Use color with consistent meaning. -Use black for all text. -Use one color (not red), font, and 3D effects for clickable items. -Use a bright color and a bigger font for titles. -Be careful with color combinations. Do not use bright colors on white.
Malfunctions Related to Dialogs					
M2	16/ User 1	When the user clicks on "Preferences", "Compound Window", a popup dialog box appears. At the bottom of the dialog box are two buttons: "Abort Changes" and "Save Changes". If the user clicks on "Abort Changes", another popup box appears, asking the user whether or not to save current changes. This popup box is redundant.	-Avoid redundancy.	4	-Do not display the second dialog box.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M3	16/ User 1	To select pathways from a list of pathways, the user used "Shift+left mouse key" to do the selection. The user did not realize that holding down the shift key is not necessary.	-The system must be clear to the user. Provide help in case of ambiguity.	5	-Provide a description on how the user can select menu items on the dialog itself.
M4	17/ User 3	BUG: When the user clicks on "Display Expression Data", then selects "File", then "Retrieve Color Scheme Parameters", then "Cancel", the application is jammed.	-The system should be thoroughly tested, and as much bug-free as possible.	2	-Fix the bug.
M5	18/ User 3, 4	The popup box produced when the user clicked on "Highlight", "Class", "amino acid biosynthesis", "All" was located at the bottom of the screen and was half hidden. The user did not realize that there are buttons at the bottom, until the instructor told the user to drag the box to the top.	-When dialog boxes are popped out, they must be visible to the user.	4	-Make the position of the popup box in the middle of the screen.
M6	19/ User 1, 2, 4, 5	Ambiguity of the "Any" and "All" items in the combo box in the "Species Comparison" dialog box. Some users selected all species without changing the combo box to "All". Others changed the combo box to "All" without selecting any species. In the later case, clicking "Ok" button did not do anything. The "Ok" button should have been disabled.	-Disable items that are not valid in the current context. -Avoid redundancy.	3	-When the user selects "All" in the combo box, automatically select all organisms, so that the user does not have to select all species in the list. -Disable the "Ok" button until at least one item is selected in the list.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Feedback and Error Handling					
M7	12/ User 1	After selecting all carbon compounds degradation pathways, the results panel states "Inserted 25 items on Answer List", which misleads the user into thinking that there are 25 pathways, not 26.	-Provide more meaningful feedback to user.	4	-The results panel should state "There are 26 carbon compounds degradation pathways. Press "Next Answer" to view the next pathway".
M8	12/ User 5	When the user uses the main window to search for a pathway, by selecting pathways from the "Organisms' summary page", a list of clickable pathways is shown. The user did not realize that the number in brackets besides the pathway class refer to the number of pathways in the class.	-Provide more meaningful feedback to user.	5	-Inside the brackets, state "Number of pathways: #", instead of just stating the number, so that the user knows what the number is for.
M9	13/ User 1	Some error messages are not too clear the user. When the user clicks on "Pathway Mode", "Get Pathway by Name", and types in "arginine biosynthesis", the following error message is displayed: "arginine biosynthesis is not a know child of reaction. Try again. (type quit to abort)". This error message is not too clear in terms of the conceptual model and does not explain why the error occurred and possible solutions.	-State error messages in terms of the conceptual model. -Provide as much detail as you can about errors. -Display causes and alternative solutions, if any.	4	-Provide error messages in red big font on the dialog box itself and point the cursor to the errant item. -Make the error message more descriptive. One possibility is "Could not find the "arginine biosynthesis" pathway in E. coli. Did you mean "arginine biosynthesis II"?".

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M10	15/ User 1 12/ User 2, 3, 4, 5	The user experienced difficulties in going back to a previous state. To go back, the user has to click on the "Backward in History" item in the left hand side menu. When the user did not find any link or arrow on the main window, the user thought that the application does not allow rolling back to a previous state. All users were given a hint that there is a way to go back.	-Keep labels unambiguous. -Frequently used items must be placed in an obvious location.	3	-Provide a tool bar at the top, and just underneath the menu. -Provide 2 icons in the toolbar, one with a left arrow and the word "Back" to go backward in history and one with a right arrow and the word "Forward" to go forward in history.
Malfunctions Related to Font					
M11	18/ User 2 12/ User 3	Error messages were not noticeable to the user. Error messages are displayed in small font in the results panel, located near the bottom of the screen, just on top of the help area.	-Use color, font, and special effects to draw the attention of the user.	3	-Use bigger font and red color to display errors. -Produce error message right where they occur.
Malfunctions Related to General Category					
M12	19/ User 1 17/ User 2	Difficulty recalling. User commented in step 19 "I saw shared somewhere. I don't know where". In step 17, user 2 remembered seeing "Overview" somewhere and could not remember where.	-The system must be clear to the user. -The system must provide minimal cognitive efforts.	4	-A better organization of the items in the tool would allow for better recalling.
Malfunctions Related to Menu					
M13	12/ User 1, 3, 4	User confused "Compound Mode" and "Pathway Mode".	-Keep labels unambiguous	4	-Change "Compound Mode" to "Search for compound(s)". -Change "Pathway Mode" to "Search for pathway(s)".

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M14	12/ User 1 18/ User 2	The way menus are organized does not allow the user to pick from higher levels directly. Clicking on a menu produces another popup menu and so on until the end menu. The user commented that it's annoying how menus keep popping up. In addition, since the syntax for selecting a menu item from the popup menu is "click-position-click", it is not intuitive on how to get rid of the menu, which is by clicking outside the menu. The user also did not know how to resize it. The user found this syntax very.	-Keep menus self-explanatory - Provide visible feedback (color, font, 3D, etc) so users know where they are in a modal menu hierarchy. -Allow users to pick from higher levels directly. -Provide a clear and easy way out of each state	4	-Use a "press-drag-release" menu. It's more intuitive and familiar to the user. -Follow UI guidelines in column 4.
M15	13/ User 1	The user was annoyed that typing part of the pathway name in the "Get Pathway by Name" does not work.	-Allow flexibility in entering information.	4	-Allow the user to type in partial pathway names (or keywords in the pathway name).
M16	13/ User 1 16/ User 3	The "Get Pathway by Name" menu item is the same as the "Get Pathway by Substring" item, except that the latter is more flexible in the sense that a portion of the pathway can be typed.	-Avoid unnecessary redundancies.	5	-Remove the "Get Pathway by Substring" item and make the "Get Pathway by Name" item more flexible (See M2).
M17	14/ All users	The user did not realize that right clicking on a clickable item brings up a menu of options. This is because the user tried right clicking outside clickable items and it did not work.	-The system must be intuitive.	4	-Using special effects and color to show that an item is clickable solves this problem.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M18	16/ User 1, 3, 4	The command menu item "Overview Mode" caused some confusion to the user. The user thought that the "Overview Mode" might allow the user to show pathways simultaneously.	-Keep labels unambiguous	4	-Rename the menu to "Graphical Overview of All Pathways".
M19	16/ User 1, 2	The menu item "Fix Window" was ambiguous to the user. The user thought that "Fix Window" would fix the diagram and that the user can search for other pathways and go back to the diagram.	-Keep labels unambiguous	4	-Remove the "Fix Window" menu item. It does not seem to offer any functionality that the user may need. -If "Fix Window" is needed, then provide help.
M20	16/ User 1	Ambiguity between the "Preferences" and "Special" menu items and between the items in their submenus. The user did not find the labels of these menus and submenus intuitive. At one point, the user commented that "special means nothing to me".	-Keep labels unambiguous	4	-Change "Preferences" label to "Format". -Remove the "Special" menu or give it a more meaningful name, like "Tools" or something that is indicative of the items it contains.
M21	12, 16/ User 1	The user experienced difficulties navigating through the returned results. "Next Answer" wasn't too clear to the user.	-The system must be intuitive.	4	-Allow navigation directly on the main window, by providing arrows (links) to go to the next or previous pathway.
M22	16/ User 2	If the user clicks on "Fix Window", then on "Backward in History", a clone of the main window is produced and the software is jammed. The application must be restarted.	-Provide a clear and easy way out of each state. -Provide details about errors.	2	-Fix the bug.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M23	16/ User 3	The "Backward in History", "Forward in History" and "Select from History" items of the cloned window should be disabled, until the user navigates through the main window.	-Disable items that are not valid in the current context.	4	-Disable the items, until the user navigates through the cloned window.
M24	16/ User 4	Ambiguity of the "Show frame in other species", "Show pathway in overview" and "Display pathway information in new window" menu items, produced by right clicking on a pathway name. The user thought that the later would bring out the pathway in a separate window, which was not the case. The user thought the labels are ambiguous.	-Keep labels unambiguous.	4	-Change the "Show frame in other species" title to "Change species". -Change the "Show pathway in overview" to "Highlight current pathway in the overview". -The "Display pathway information in new window" should display the pathway in a new window.
M25	17/ All users	Ambiguity of the "Show Key" command of the "Overview Mode", which shows the legend. This command should have been placed in the main menu.	-Keep labels unambiguous	4	-Make the label of link "Legend" instead of "Show Key". -Provide a link to the legend in the main menu.
M26	17/ User 1	The commands menu does not provide vertical scrolling capabilities. Thus, if more items are added to the menu, as a result of clicking on a mode, than the capacity of the space allocated for additional items, such as in the case of "Overview Mode", those items are added on the top of the commands menu.	-Provide scrolling when there are more items than what fits in the panel.	4	-Provide vertical scrolling. -Do not add the items at the top, because they may block other items.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M27	17/ User 3	The user thought that "Display Expression Data" would allow the user to show a legend.	-Keep labels unambiguous	4	-Provide a help on this item.
M28	18/ User 1 19/ User 2	Ambiguity of the "By Class" and "All by Class" labels, produced in a menu by clicking on "Overview Mode", then "Highlight", then "Find and highlight" then "Pathway". The user thought that "All by Class" means all items of a pathway class. But it actually means highlighting all pathways in the overview according to class.	-Keep labels unambiguous	4	-Change the label of "by Class" to "Class…". The three dots indicate that a popup dialog box or menu will appear. -Change the label of "All by Class" to "All Pathways by their associated class".
M29	18/ User 2 12/ User 3	Ambiguity between "Get Pathway by Name" and "Get Pathway by Class". The user misused these menu items.	-Keep labels unambiguous	3	-Have one menu for searching for pathways, titled "Search for pathway(s)…". Clicking on it would produce a popup box, which has a field to allow typing a partial pathway name, with the names of matched pathways appearing in the list box underneath. Provide the same functionality for searching for pathways by class.
M30	18/ User 5	"Restore highlights from file" is ambiguous. The user clicked on it when asked to highlight pathways.	-Keep labels unambiguous	4	-Change "Restore" to "Import" because it's more commonly used. -Provide help on the item.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M31	19/ User 2	The user did not notice the "Species Comparison" item underneath the "Highlight" menu of the "Overview Mode". This is because "Species Comparison" is inconsistent with the rest of the items under the "Highlight" menu.	-Be consistent in grammar and pattern in sets of menu labels.	4	-Place the "Species Comparison" as a sub-item of the "Pathway" item and have another item "Class" in the pathway to highlight pathway by class. Rename "Species Comparison" to "Shared with Species". -Change labels "Pathway" and "Gene" to "Pathway(s)" and "Gene(s)" to be consistent.
Malfunctions Related to Response Time					
M32	16/ All users	Loading up the pathways on the "Overview Mode" takes time, and the user cannot do anything until the whole page is loaded up. On an Intel Pentium® 4 CPU 1.60GHz, it takes about 14 seconds. The first time the user clicked on "Overview Mode", the user was fascinated by the graph. However, in the subsequent times, the user was annoyed, especially when clicking on it by mistake.	-Provide a 'cancel' mechanism for operations in progress.	3	-Provide a "Cancel" mechanism. -Provide a time bar, showing how much time is needed for the whole page to load up.
Malfunctions Related to Windowing Interface					
M33	15/ All users	Ambiguity of the '+' and '-' items. The user did not realize that '+' and '-' items represent activators and inhibitors, respectively.	-Use special 3D effects to draw attention of user.	4	-Enlarge the '+' and '-' items and use the same color and font as clickable items.

VM#	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
M34	15/ User 1 12/ User 3	The results panel had "Command:" printed in it, misleading the user into think that the use of this area is to type in commands. The user tried typing inside that panel to do searching or to go back.	-Avoid ambiguity.	3	-Enlarge the results panel. -Make the font of messages in the results panel bigger. -Remove the word "Command:" from the results panel to avoid misleading the user into thinking that this panel is for typing commands.

Table B-9: A list of videotaped evaluation malfunctions (VM) in WIT, including task # that uncovered the malfunction (See Appendix B.3), user #, malfunction description, UI guidelines violated, severity level from 1 to 5 (See Table 3-1), and recommendations for change

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Feedback and Error Handling					
W1	28, 30/ All users	The search results displayed are not intuitive. The results are shown as a list of non-meaningful names. All users expressed confusion as to what the results mean.	-Keep names unambiguous. -Provide tool tips or balloon help text for entry.	3	-Provide meaningful names in the search results and a description of the name. -Provide a short description on what each link provides.
W2	28/ User 1, 4	The user looked confused when clicking on first entry of the pathway search results, which displayed tables of information, with no pathway diagram. To display a diagram, the user has to click on the "Diagram Picture" link.	-Avoid ambiguity.	4	-Since the pathway tool is responsible for showing pathways, perhaps a better way is to display a pathway diagram right away, with additional information and other links underneath.

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
W3	28/ User 2	Clicking on substrates in the diagram did not display information.	-Provide feedback to user. -Provide expected functionality.	4	-Fix the problem of why information cannot be displayed.
Malfunctions Related to Information Query Interfaces					
W4	28/ User 1	Ambiguity of the "Query Pathway" link. The user commented that simple words, such as "search", are better than "Query".	-Avoid ambiguity.	4	-Change "Query Pathway" to "Search for Pathway(s)".
W5	28/ User 1, 2	Cannot search pathway by name or by class. This tool only allows one to search pathways containing compound(s). The search confused the user a little.	-Provide searching flexibility.	4	-Make the search more flexible. -Allow users to search pathways by name.
W6	28/ User 1	Ambiguity of the "Check Data" link. The user thought that "Check Data" may lead to linked pathways.	-Keep labels unambiguous. -Provide tool tips or balloon help text for each field.	4	-Provide help (tool tips) on the use of the "Check Data" button.
W7	28/ User 2	On the main WIT page, there are some search links under the "Screen shot" title, which show you what the commercial WIT tool search forms look like. These links made the user think these can actually be used for searching.	-Avoid ambiguity.	5	-Perhaps the best way is to make "Screen shot" a link that takes you to a different page, with all of the links to search sample screen shots.
W8	28/ User 2	Typing an organism name in the "taxon" field does not work. No help was provided.	-Provide tool tips or help for each field.	3	-Provide help (tool tips) on how to enter information in the "taxon" field.

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
W9	28/ User 2, 3, 4	Ambiguity of the "In all volumes" combo box.	-Keep labels unambiguous. -Provide tool tips or help for each field.	4	-Provide help (tool tips) on the "In all volumes" combo box.
W10	29/ User 1, 2, 4, 5	Enzymes represented by their E.C. number only on diagrams, which makes it hard to locate enzymes on the figure, unless the user knows the reaction or the E.C. number.	-Avoid ambiguity.	3	-Provide the enzyme name. Provide the E.C. number when the user clicks on the enzyme for more information.
W11	29/ User 1	Ambiguity of the "R" item on the pathway figure. The figure should have a legend.	-Keep labels unambiguous. -Provide tool tips or help.	4	-Provide help (tool tips) on the "R".
W12	30/ All users	The enzyme information page was very confusing to the user. 3-letter abbreviations were used and users commented that these abbreviations are ambiguous. To know what the abbreviations mean, the user needs to click on them. The page is also not nicely organized. Users commented that it's hard to find information.	-Avoid ambiguity.	4	-Do not use abbreviations, especially when these abbreviations are not standards. Use the full name. -Provide help (tool tips) on the links. -Reorganize the information, such that the most important items are at the top.
W13	30/ User 1	Spelling mistake on the results page. The word "systematic" should have been "systematic".	-There should be no grammar or spelling mistakes.	5	-Change "systematic" to "systematic".
Malfunctions Related to Response Time					

VM #	Stage/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
W14	28/ All users 30/ User 1, 2	The website was extremely slow (perhaps because it's a government website). The user was annoyed that it was taking a long time.	-Announce long delays (> 3s). -The user will tolerate long delays (up to 15 seconds) only when loading up complex queries. A time longer than that is intolerable.	2	-Research the reason behind the long delays. -Provide a description of why the delays occur and how long it will take to display information.

Table B-10: A list of videotaped evaluation malfunctions (VM) in BioPathVis, including task # that uncovered the malfunction (See Appendix B.3), user #, malfunction description, UI guidelines violated, severity level from 1 to 5 (See Table 3-1), and recommendations for change

VM#	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
Malfunctions Related to Color					
BP1	3, c, ii / User 1, 2, 3	Selecting an item in the "Search in graph" combo box takes the user to the item and highlights it. However, the highlighting is done in a green dotted line, which is not very visible.	-Use bright colors to draw the attention of the user.	4	-Highlight the compound using a red solid line.
Malfunctions Related to Dialogs					

VM#	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
BP2	5 / User 1	Ambiguity of the "Select pathway name(s) to search" item. Underneath that label is a text field, followed by a list box. Typing a name in the text field takes the user to the pathway that matches that name in the pathway list.	-Keep labels unambiguous.	4	-Have a more descriptive label. An example is "Select pathway name(s) to search. Entering a pathway name in the text field below takes you to the matching pathway name in the list".
BP3	5 / User 2, 3	Although user figured out that holding down 'Ctrl' selects multiple pathway names, users 2 and 3 thought it would be better to state that somewhere.	-The system must be clear to the user.	5	-Provide a tool tip that tells the user about the use of 'Ctrl' and 'Shift' for multiple selections.
BP4	5 / User 2, 4	The user was confused a little about the search results being displayed as a tab next to the quick search tab. The search results tab has the same structure as the quick search tab, but it consists solely of the selections the user made using the advanced search dialog box. When asked further, the user thought that this is a good place to view search results, but that it's confusing a little at the beginning.	-The system must be intuitive.	5	-Provide a tool tip that tells the user where the search results will be displayed upon clicking on the "Ok" button.
BP5	6 / All users	When asked to view all amino acids and lipids, the user did not click on the "Select All" item. Items should have been selected by default.	-The system must be intuitive.	5	-Select all items by default.
Malfunctions Related to Help					

VM#	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
BP6	2, a; 3, b / User 5	The user considered the help menu to accomplish certain tasks and the user did not find anything there.	-Provide help when necessary.	4	-Even though tool tips are provided, information must be provided in the Help menu as well.
Malfunctions Related to Pathway Graphs / Overview Images					
BP7	2, a / All users	The user first tried to single click on the overview image of carbohydrate metabolism. The user first thought that single clicking would work, but then realized the item had to be double clicked.	-The system must be clear to the user. Provide help in case of ambiguity.	5	-Even though a tool tip was provided to double click, the user did not look at the tool tip. This ambiguity was due to the fact that overview images are static images, whereas the actual pathways are dynamic. Making the overview images dynamic would solve this problem.
BP8	3, c, ii / User 1, 3	Scrolling of the graph, using the mouse middle button, is not fast.	-Provide the expected functionality.	4	-Provide quicker scrolling.
BP9	3, c, iii / User 1, 2, 3	Double clicking on the reaction node to display information is very sensitive. The user has to double click right inside the small reaction node. Part of this problem has to do with the fact that edges are selectable. Hence, when the user double clicks, if the user moves the mouse a little, one click is done on the reaction node itself and the other on the edge.	-The system must be clear to the user.	4	-Make the edges non-selectable.
Malfunctions Related to Toolbar and Menu					

VM#	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
BP10	3, a, i / All users	Hiding the legend is far from the actual legend. Hiding the legend is provided both in the toolbar and in the "View" menu. When users tried to hide legend, the user first tried clicking on legend before looking at the toolbar or "View" menu.	-The system must be intuitive.	5	-Provide a tool tip next to the legend. Consider allowing the user to hide the legend by clicking on it, perhaps by making the legend a panel as opposed to a toolbar with an 'x' on the top to hide it.
BP11	3, c, ii / User 1, 2, 3	The legend does not have all info related to pathway graphs. User 1 looked for the enzyme representation in the legend and could not find it. User 2 looked for the reaction representation the legend and could not find it.	-Provide all necessary information.	4	-Provide all necessary information in the legend.
BP12	3, c, ii / User 1	To search for a compound in a pathway graph, user 1 tried double clicking on the "Search in graph" label in the toolbar on top of the pathway graph, as opposed to searching for compounds in the combo box beside it. The user later figured out that the user has to use the combo box.	-The system must be intuitive.	4	- Improve the relationship between the label and the combo box by changing the background color of the combo box to white as opposed to gray.
BP13	3, a / User 3	The legend is a floatable toolbar and it's located at the bottom of the pathway quick search panel. Once the legend is moved to the side of the pathway quick search panel, it's hard to place it back at the bottom. However, it could be easily placed at the top or the side.	-Provide the expected functionality.	5	-Change the floatable functionality to allow the legend to be easily dropped anywhere the user would like it to be.

VM#	Task/ User #	Problem Description	UI Guidelines Violated [21]	Severity Level	Recommendation for Change
BP14	3, c, i, v / User 5	When asked to expand the pathway graph panel, user 5 considered the "View" menu and did not find that capability there. When asked to hide co-substrates and enzymes, user 5 also considered the "View" menu, since hiding a legend and zooming in and out was provided in the "View" menu. The user later considered the toolbar on top of the pathway graph to accomplish these tasks.	-Provide the expected functionality. -Provide consistency.	4	-Provide an option in the "View" menu to expand the whole pathway or to hide the search panel. -Provide an option in the "View" menu to hide enzymes, co-substrates, etc from the pathway graph.

B.6 BioPathVis Follow up Questionnaire

The follow up questionnaire for BioPathVis will be placed on the web site of this thesis: http://www.site.uottawa.ca/~tcl/gradtheses/rkhartab.

Appendix C Questionnaire

The questionnaire for collecting data on user preferences will be placed on the web site of this thesis: http://www.site.uottawa.ca/~tcl/gradtheses/rkhartab.

C.1 Malfunction Data

Table C-1: A list of questionnaire malfunctions (QM) for BioCarta, found by users' responses in questionnaires, including user #, malfunction description, UI guidelines violated, related videotaped evaluation malfunction (VM – See Table B-5), severity level from 1 to 5 (See Table 3-1), and recommendations for change

QM #	User #	Problem Description	UI Guidelines Violated [21]	Related VM #	Severity Level	Recommendation for Change
QB1	User 1, 2, 8, 11	Information on certain enzymes is not present. The user commented that one "Could not find the specific carbamoyl-phosphate synthase enzyme on 'search' or 'Query'", because only the synthetase is present, and not the synthase. Also, user 2 commented that you "cannot click on anything to gain more information", since nothing on the figure, except enzymes, is clickable.	-Provide the necessary functionality and information.	-	4	-Provide the necessary information.
QB2	User 2	There is a legend, but certain molecules are not identified.	-Provide help on items.	-	4	-Provide a legend.
QB3	User 4	No links between pathways.	-Provide the necessary functionality.	-	4	-Show links to other pathways.

Table C-2: A list of questionnaire malfunctions (QM) for ExPASy, found by users' responses in questionnaires, including user #, malfunction description, UI guidelines violated, related videotaped evaluation malfunction (VM – See Table B-6), severity level from 1 to 5 (See Table 3-1), and recommendations for change

QM #	User #	Problem Description	UI Guidelines Violated [21]	Related VM #	Severity Level	Recommendation for Change
QE1	User 2, 3, 4, 8	Pathway figures are confusing to look at and are cluttered. User 3 commented that the graphs "should be simplified". User 8 commented that it's "too complex for teaching".	-Present information in a way that is easy to understand.	-	5	-Modify the figure to make it less cluttered and easier to look at.
QE2	User 2, 4	Absence of legend.	-If help is necessary, provide it.	E6	3	-Provide a clear legend of what each color and shape in the figure means.
QE3	User 6	Enzyme search brings the user to a list of pathways, but the user doesn't know which pathway to click on.	-Keep labels unambiguous.	E10	4	-Provide a more meaningful name, or provide a description of the name.
QE4	User 6, 12	It is not easy to move from one pathway to another. Since the pathway maps are arranged in grids, one might have to click several times to get to the desired information.	-Allow the user to get a clear picture of what the user is looking for.	E7	5	-Allow the user to display more than one map at a time, because this will make it easier for the user to see all the pathways surrounding a certain compound.
QE5	User 11	Some regulatory pathways are not up to date.	-Provide correct information.	-	4	-Update all regulatory pathway information.

Table C-3: A list of questionnaire malfunctions (QM) for KEGG, found by users' responses in questionnaires, including user #, malfunction description, UI guidelines violated, related videotaped evaluation malfunction (VM – See Table B-7), severity level from 1 to 5 (See Table 3-1), and recommendations for change

QM #	User #	Problem Description	UI Guidelines Violated [21]	Related VM #	Severity Level	Recommendation for Change
QK1	User 1, 4	Absence of legend on diagrams. The help link was all the way to the bottom of the main page.	-Provide help.	K10	4	-Provide a legend right on the figure itself.
QK2	User 2	Enzymes have to be known by E.C. number.	-Avoid ambiguity.	K8	3	-Provide the enzyme name and the E.C. number. Provide the name in the figure and the E.C. number when the user clicks on the enzyme.
QK3	User 2	Absence of enzyme 3D structures on the pathway figure.	-Provide the necessary information.	-	4	-Provide the option of having chemical structures.
QK4	User 11	The diagrams are incomplete (e.g. co-products are not shown).	-Provide the necessary information.	-	4	-Provide the necessary information.

Table C-4: A list of questionnaire malfunctions (QM) for MetaCyc, found by users' responses in questionnaires, including user #, malfunction description, UI guidelines violated, related videotaped evaluation malfunction (VM – See Table B-8), severity level from 1 to 5 (See Table 3-1), and recommendations for change

QM #	User #	Problem Description	UI Guidelines Violated [21]	Related VM #	Severity Level	Recommendation for Change
QM1	User 1, 7	Too much cognitive work. Information is difficult to search.	-The system must provide minimal cognitive efforts.	M1, M10, M12, M13, M21, M29	4	-A better organization of the items in the tool would allow for better recalling and searching.
QM2	User 1, 9, 10	Absence of legend on diagrams.	-Provide help.	M25	4	-Provide a legend.
QM3	User 2	Confusing command structure.	-Avoid ambiguity.	M34	3	-Remove the word "Command:" from the results panel to avoid misleading the user into thinking that this panel is for typing commands. -Move the results panel to the top of the main window, and just underneath the toolbar.
QM4	User 3	When the user clicks on the '+' and '-', not all names of activators and inhibitors are printed in the results panel. Clicking on the enzyme, however, provides all of that information	-Provide consistent information.	-	5	-Provide all the activators and inhibitors when the user clicks on the '+'/'-'.

QM5	User 4	The tool froze a couple of times.	-See M22 and M4.	M22, M4	2	-See M22 and M4.
QM6	User 9, 10	Poor menu design.	-See M13- to M31	M13 to M31	4	-See M13- to M31
QM7	User 9, 10	No rollback capabilities.	-See M10	M10	3	-See M10
QM8	User 9, 10	Too much color.	-See M1	M1	4	-See M1

Table C-5: A list of questionnaire malfunctions (QM) for WIT, found by users' responses in questionnaires, including user #, malfunction description, UI guidelines violated, related videotaped evaluation malfunction (VM – See Table B-9), severity level from 1 to 5 (See Table 3-1), and recommendations for change

QM #	User #	Problem Description	UI Guidelines Violated [21]	Related VM #	Severity Level	Recommendation for Change
QW1	User 1, 2, 7, 4	The website is too slow to load. The website experiences technical difficulties	-Announce long delays (> 3s). -The user will tolerate long delays (up to 15 seconds) only when loading up complex queries. A time longer than that is intolerable.	W14	2	-Research the reason behind the long delays. -Provide a description of why the delays occur and how long it will take to display information to the user.
QW2	User 4	The tool is confusing and not very clear.	-See W1, W2, W4, and W12.	W1, W2, W4, W12	-See W1, W2, W4, and W12.	-See W1, W2, W4, and W12.
QW3	User 4	Too many short forms and links on the enzyme information page.	-See W1 and W12.	W1, W12	-See W1 and W12.	-See W1 and W12.

Appendix D Introduction to Statistical Analysis Techniques

This appendix is devoted to reviewing two main statistical analysis techniques that are used in this research. These two techniques are the ANOVA test and the T-Test.

D.1 ANOVA test

When dealing with averages involved with more than two population means, we need to conduct an analysis of variance (ANOVA) test to determine whether there are significant differences between the means of the populations, or whether the differences were purely due to random chance.

Why do we need the ANOVA test? To compare three means, we need three tests to compare each mean to the other. Each test is subject to the possibility of error. To compare four means, we need six tests. The more tests we perform on a set of measurements, the more likely it is that at least one of our conclusions will be incorrect [25]. The ANOVA test provides one overall test to judge the quality of the population means [25]. Once we have determined whether there is actually a difference in the means, we can use the T-Test to find out where the differences lie [25].

In the ANOVA test, we will assume that the samples are randomly and independently selected from their respective populations and that the populations are normally distributed with equal means $\mu_1, \mu_2, ..., \mu_k$ and equal variances $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$ [25].

Let x_{ij} be the *j*th measurement in the *i*th sample. We first consider the total variation in the experiment, which is measured by a quantity called the total sum of squares (SS). The total SS is partitioned into two components, called the sum of squares for treatments (SST), which

measures the variation among the k sample means, and the sum of squares for error (SSE), which is used to measure the pooled variation within the k samples.

$Total\ SS = SST + SSE$ [25]

Where

$SST = \sum \frac{T_i^2}{n_i} - CM$; The correction of the mean $CM = \frac{(\sum X_{ij})^2}{n} = \frac{G^2}{n}$; G represents the grand total of all n observations [25]

And

$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + ... + (n_k - 1)s_k^2$ [25]

Each of the sources of variation, when divided by its appropriate degrees of freedom, provides an estimate of the variation in the experiment [25]. Since Total SS involves n squared observations, its degrees of freedom are $df = (n-1)$ [25]. Similarly, SST involves k squared observations, and its degrees of freedom are $df = (k-1)$ [25]. Finally, the sum of squares for error has $df = (n_1 - 1) + (n_2 - 1) + ... + (n_k - 1) = n - k$ [25]. These two sources of variation and their respective degrees of freedom are combined to form the mean squares: $MS = \frac{SS}{df}$ [25].

The mean squares in the ANOVA table can be used to test the null hypothesis (H_o) that $\mu_1, = \mu_2 = ..., = \mu_k$, versus the alternative hypothesis (H_a) that at least one of the means is different from the others [25]. We reject H_o if $F = \frac{MST}{MSE} > F_\alpha$ where F_α lies in the upper tail of the F distribution [25].

D.2 T- Test

When dealing with averages involved with two population means, we need to conduct a T-Test between each pair of tools. We will consider 2 populations at a time, having means μ_1 and μ_2 and variances σ_1^2 and σ_2^2. The T-Test can be applied when the samples within the populations are small and the population variances are unknown, provided that both populations follow the normal distribution, which is the case here [29]. Under these conditions, the sampling distribution of the statistic

$t = \frac{(\bar{x}_1 + \bar{x}_2) - \delta}{\sqrt{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})}}$ (Where δ is a specified constant;, $\bar{x}_1$ and $\bar{x}_2$ are sample means;, n_1 and n_2 are the sample sizes;, and s_1^2 and s_2^2 are the sample variances) has a t distribution with $n_1 + n_2 - 2$ degrees of freedom [25].

As such, we will test the null hypothesis (H_o) that $\mu_1 - \mu_2 = \delta$, and we will choose zero as the value of δ [29]. Accepting the null hypothesis means that there is no significant difference between the means of the two populations, and that the differences were purely due to random chance.