Visualisation of the information resources for cell biology

Andrei Malchanau, Anton Nijholt and Hans E. Roosendaal

University of Twente

HMI Research Group

Post box 217

7500 AE Enschede, the Netherlands

Email: A.V.Malchanau@ewi.utwente.nl

ABSTRACT
Intelligent multimodal interfaces can facilitate scientists in utilising available information resources. Combining scientific visualisations with interactive and intelligent tools can help create a "habitable" information space. Development of such tools remains largely iterative. We discuss an ongoing implementation of intelligent interactive visualisation of information resources in cell biology.

Categories and Subject Descriptors
H.5.0 [**Information Systems**]: Information Interfaces and Presentation – General.

General Terms
Design, Experimentation, Human Factors

Keywords
Visualisations, Interfaces, Knowledge representation, Information systems

1. INTRODUCTION
Multimodal interaction can improve communication between human and computer system. Multimodality often implies fusion of several modalities. Interpreting multimodal input may require understanding of the information that is being communicated. Communicating with variety of information resources needs intelligent multimodal interfaces.

There are thousands of information resources in biology alone. Many resources are available for scientists on Internet directly or through institutional portals such as European Bioinformatics Institute (EBI) or National Center for Biotechnology Information (NCBI) websites. Elsewhere [5] we have put forward the concept of Habitable Interfaces as guidance in developing interfaces to the content of scientific archives. The concept is based on a model of scientific communication. The model points towards domain knowledge as a means to structure information from various information resources. This underlines the need for intelligent multimodal interfaces.

We are developing technologies for intelligent interactive visualisations of information resources as a part of Habitable

Interface. Evaluation of such technologies needs an implementation. Here we discuss a prototype of Habitable Interfaces for cell biology and, partially, the design of the evaluation of the prototype.

In the development of the prototype we aim at the existing information resources in biology. These resources include publicly available and proprietary databases and a variety of tools. A comprehensive overview of publicly available databases and tools can be found in [8].

Traditionally, scientists employ a variety of visualisation means such as charts, maps, tables and graphs to cope with the complexity and the vast volumes of information in fields like cell biology. Some information resources on Internet offer interactive versions of traditional or new visualisations. Visualisations can be of different types. Some of them are map-like others are diagrams. Clearly, large maps require zooming and panning. Diagrams, on the other hand, can be easier to develop and to support interactivity. We are primarily interested in visualisations that are familiar and useful for the biologists to interact with the content of digital archives.

2. VISUALISING THE INFORMATION SPACE
Visualising the information space needs some general model to map the information resources. For some domains like cell biology finding the general model (such as a biological cell) can be rather straightforward. Other domains may require more abstract models that can be expressed as tables or graphs.

Visualisation and interaction with the models should be described in some language. In general a description may include visual representation, navigation (such as zooming into components of a general model) and the meaning of the visualisation in terms of a particular application.

Browsing through visualisations by zooming into parts of a model like a biological cell usually requires deciding what particular meaning of a part selected by the user is intended. For example, zooming into the mitochondrion of a cell may require visualising the inner structure of the mitochondrion or, alternatively, the biological processes that take place in mitochondria depending on the user intention. In some cases deciding what visualisation is required can be derived from the previous steps in browsing, in other cases such a decision may need to be taken by the user. In the last case a set of relevant visualisations is proposed to the user.

Clearly, to enable the interaction with these diagrams at the level suitable for biologists the underlying technological system should be able to "reason" about maps presented on a screen. For example, zooming within a visualisation may require emphasising some elements while removing irrelevant details. Also some components of a biological cell can be related to hundreds of pathways (mitochondrion in Figure 2 relates to many of the pathways in Figure 1. Both figures are reproduced from [1] to give some orientation about traditional visualisations in biology). This requires some means of representing the pathways and selecting those that are of interest for a biologist.
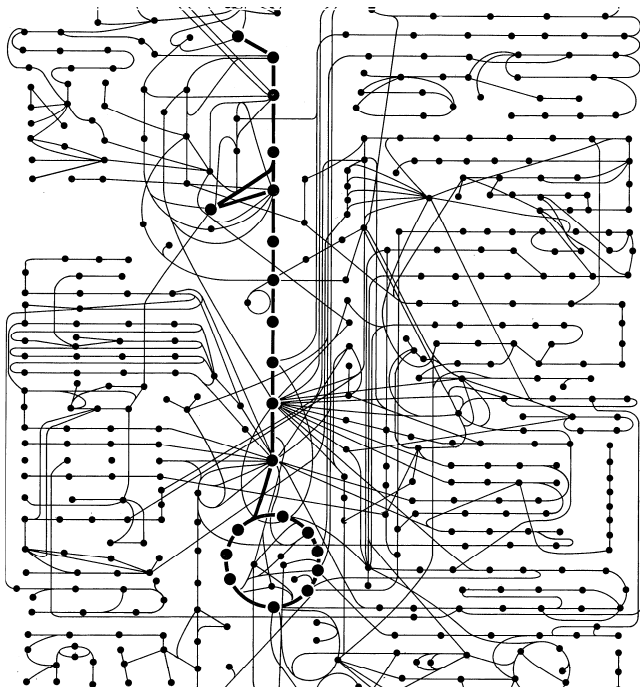

Figure 1. A biological diagram from a textbook representing some of the metabolic pathways (about 500) in a typical cell. Filled circles represent molecules.

We distinguish between the following tasks for reasoning about elements of visualisation:

- Distinguishing between different concepts presented on a map

- Scaling

    o Displacement of elements

    o Symbolization

    o Emphasizing

- Mapping data

    o Placing symbols

    o Gradient fill and histograms

    o Organizing "nice" layout

The first task is rather simple and requires naming parts of the images so that they can be referred to.

Scaling can be non-trivial. For example, scaling of maps requires that mapping of data to be scaled along with the image. At least, three subtasks can be identified within scaling. Changing scale may need the displacement of some elements and mapped data so as to keep "proper" mutual references. Symbolization can be used to abstract from details when the scale becomes large. Emphasizing refers to cases when some elements of a map that are considered important are scaled disproportional with the rest of the map. All these subtasks are interrelated. For example, emphasising some elements may require the displacement of these or other elements.
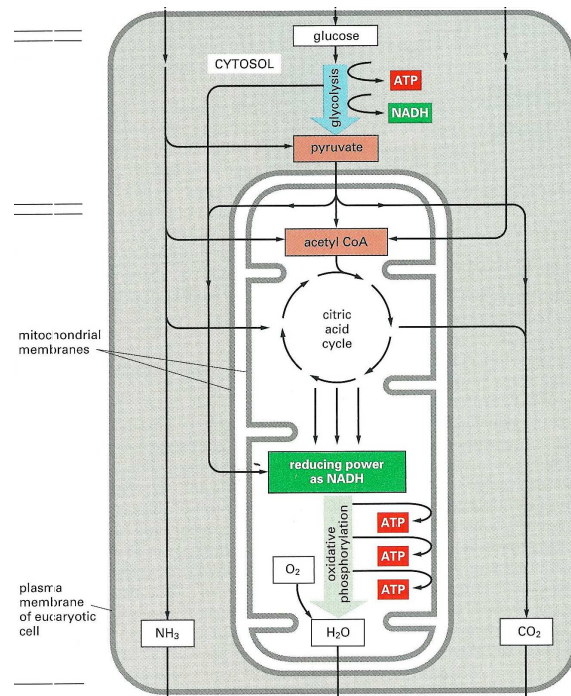

Figure 2. A part of simplified (generalised) diagram of cellular metabolism in animal cells. Processes are mapped onto cellular components (cytosol and mitohondrion)

Mapping data can be done if there is some reference from data to the visualisation that can be logically derived by a reasoning engine. Ideally, a variety of ways to map should be available for scientists. Mapping of data is rather a standard task in scientific visualisation. Criteria of "nice" layout should be defined.

The tasks enumerated above are not entirely new. Considerable progress has been made in applied research such as Geographic Information Systems (GIS) [10] and more general research in qualitative spatial reasoning [3], [9]. Cohn [2] has proposed a formalisation of the "bio-spatial" knowledge. The tasks of some QSR calculi can be understood as solving binary constraint satisfaction problem on infinite domains [7]. Despite the progress achieved, there remain some open issues in both applied and theoretical disciplines [3]. Furthermore, there are differences that our application of the maps has in comparison with GIS. Many data in GIS are geo-referenced. Such data can be directly mapped onto a pictorial representation of some terrain. In our application data may have an indirect coordinate reference as in the case of mapping diseases associated with particular genes onto a chromosome. Comparing Figure 1 and Figure 2 it is clear that the

last one cannot be derived from the first one based on data alone. Such derivation requires substantial knowledge of biology and visualisation techniques. Further, data are more likely to have reference to some concept such as nucleus as a part of the biological cell rather then a coordinate. Nevertheless some of the techniques used in GIS may be adopted in the development of Habitable Interfaces. An example of Naive Geography theory [4] can inform development of visualisations in biology. The idea of reasoning that is close to users' knowledge seems to correspond to that of Habitable Interfaces.

In the implementation we follow rather a practical approach of combining the most suitable techniques so as to minimize manual work in developing visualisations.

3. AN IMPLEMENTATION

General idea behind an implementation of habitable interfaces is presented in the Figure 3.

Both resources and Browsing interface are referring to the domain knowledge representation. User interaction is interpreted by the reasoning engine and requests are sent to the information resources.
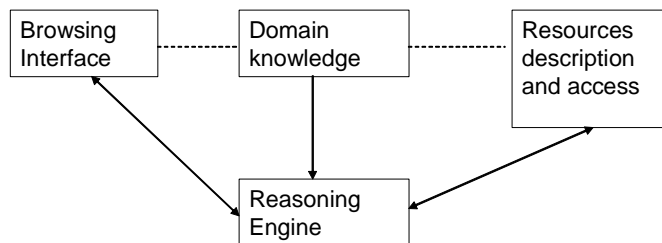
Figure 3. General idea for developing prototypes of Habitable Interfaces

A framework of the implementation is presented in the Figure 4. The prototype is implemented in the Microsoft Windows environment and aimed to experiment with the user interface visualisation features rather then networking or distributed query processing capabilities. Multithreaded environment is provided by POSIX threads (p_thread) library.

The following main elements of implementation can be distinguished:

- Language of visualisation

- Graphical visualisation

- Reasoning engine for visualisation

- Access to information resources

The last part of the implementation includes several components, but falls outside the scope of this paper.

A dialect of Prolog that includes Constraint Satisfaction module is used to reason about visualisations.

Graphical visualisation part utilises native OS (Microsoft Windows) API to produce visualisations based on the output of the Prolog engine.

We are developing a language to describe the visualisations. The language is intended to add on top of languages such as that of Scalable Vector Graphics (SVG) a layer of elements specific for the domain (cell biology). This enables automated reasoning about visualisations.

The kind of tasks in building visualisations that are possible to automate depends on the knowledge representation techniques and capabilities of the reasoning engine. Despite the progress achieved in research on knowledge representation and reasoning there is no single outstanding framework that is the best for all purposes. Consequently every application requires combination of the techniques. At present both the knowledge representation and reasoning components are being experimented with. We emphasise expressiveness and reasoning capabilities over speed of computation.
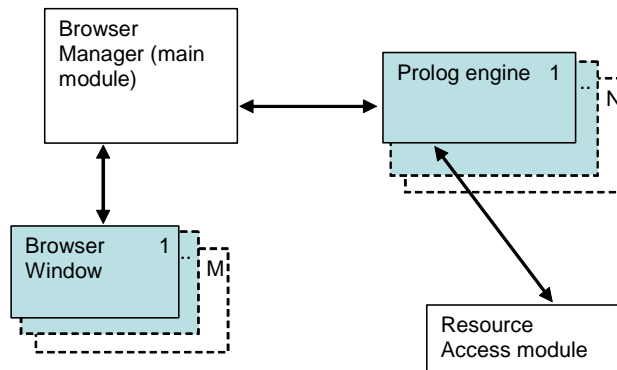
Figure 4. A particular implementation of a platform for prototyping Habitable Interfaces

The main considerations for our implementation are flexibility and ease of implementing different kinds of reasoning capabilities. The use of Prolog engine is thought to allow implementing and adjusting different types of reasoners rather quickly. Several Prolog engines can be employed at different levels of the system. These levels include processing of the zooming requests, generating visualisations, planning and optimising access to the information resources and monitoring user requests to maintain user profile.

4. EVALUATION OF THE PROTOTYPES

We plan to involve biologists as prospective users into evaluation of habitable interfaces. The background information such as experience in research and in working with information resources, position (trainee, fellow or senior researcher) will be collected from each participant prior to evaluation.

Evaluation of the prototypes is based on the following criteria:

- Data obtained in a search

- Time needed to obtain the same result

- Effort to obtain the same result

- Level of trust [6]

The measurements on the first two parts can be directly obtained from the logs of interaction of users with the interfaces. The

measurements on the last two parts of the criteria can be obtained by means of a questionnaire.

The design of evaluation that we describe here is based on a comparison of the interfaces. We plan balanced mixed design of evaluation. Each of the participants of the evaluation will compare two types of interfaces: the traditional web based interface and a prototype. Both interfaces provide access to the same set of information resources. The participants will be assigned with a task to solve by using one of the interfaces. After completing one task a participant will be proposed to complete the second task using the other interface.

Completing the tasks require knowledge of relevant information resources and availability of these resources.

The particular set of conditions (which task should be completed with which interface and in what sequence) will be chosen individually for every participant based on background (profile) information so as to balance the number of different profiles of participants assigned to every set of conditions.

After completing the assignments participants will be asked to answer a questionnaire.

Balanced mixed design of evaluation allows comparison of interfaces on the one hand and decreases interaction effects that are characteristic of within-subject designs on the other hand.

5. SUMMARY AND FUTURE WORK

We have discussed the tasks that need to be solved either by designers of interfaces or automated reasoning tools to help scientists navigating the variety of scientific information resources and tools available. A completely automatic solution of these tasks does not seem to be feasible at present. The practical solution is to minimise the amount of work designers spent on dealing with details that can be successfully automated. The plausibility of the techniques developed needs to be evaluated.

We are interested to explore interaction techniques that employ richer set of modalities such as gestures, speech and haptic feedback.

6. REFERENCES

[1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular Biology of the Cell. Garland Science, New York, 2002

[2] Cohn, A. G. Formalising Bio-Spatial Knowledge. Proc. 2nd International Conference on Formal Ontology in Information Systems (FOIS'01), edited by C. Welty and B. Smith, pp 198-209, ACM 2001.

[3] Cohn, A. G. and Hazarika, S. M. Qualitative spatial representation and reasoning: an overview. Fundamenta Informaticae 45, 1–29, 2001.

[4] Egenhofer, M. J., and Mark, D. M. Naive Geography. In Frank, A. U. and Kuhn, W., editors, Spatial Information Theory: A Theoretical Basis for GIS, Berlin: Springer-Verlag, Lecture Notes in Computer Sciences No. 988, pp. 1-15, 1995.

[5] Malchanau, A., Vet,van der, P., Roosendaal, H. Federating resources of Information Systems. Browsing Interface (FRISBI). In De Bra, P., editor, Proceedings Conferentie Informatiewetenschap, pp 85-90, 2003.

[6] Malchanau, A. Can Intelligent Environments be Trustworthy? Designing for trust in Scientific Communication: Concept of Habitable Interfaces (Position Paper), in CHI 2004 Workshop: Considering Trust in Ambient Societies, April 22, 2004

[7] Nebel, B. and Scivos, A. Formal properties of Constraint Calculi for Qualitative Spatial Reasoning. Künstliche Intelligenz, Heft 4/02, pp14-18, 2002

[8] Pevsner, J. Bioinformatics and Functional Genomics. John Wiley & Sons, Hoboken, New Jersey, 2003

[9] Stell, J. G. Part and complement: Fundamental concepts in spatial relations. Annals of Mathematics and Artificial Intelligence 0: 1–17, 2004.

[10] Stoter, J.E., Kraak, M.J., Knippers, R.A. Generalisation of framework data: a research agenda. ICA Workshop on "Generalisation and Multiple representation" 20-21 August, Leicester, 2004.