# Cytokine Information System and Pathway Visualization

**Tan Shengyang**[1] **Kwoh Chee Keong**[2]
[1]*School of Computer Engineering, Nanyang Technological University, Singapore*
[2] *School of Computer Engineering, Nanyang Technological University, Singapore*
*Email : s8133992i@ntu.edu.sg, asckkwoh@ntu.edu.sg*

**ABSTRACT**: In this paper, we highlight the development of a web application system. Its main objective is to provide pathway visualization functionalities for inter-cytokine relationships, as well as for other types of relationships, with a specific cytokine(s) of interest. A natural language processor is first used to extract information from a certain web page that concerns the cytokine(s) of interest. The results obtained are then further processed and then displayed graphically to the user. The system displays how the cytokine(s) of interest interacts with other cytokines and cells. Useful information such as the type of reaction and catalyst involved, if any, are also displayed. In addition, the system also offers functionalities for graphical manipulations of the visualized pathways. The system has been shown to provide better overview, and hence, improved learning to readers who are new to this field by virtue of accurate inputs obtained from the natural language processing module.

## 1 MOTIVATION

Since the discovery of the first cytokine in 1976, a whole range of immunologically active cytokines have been discovered. These specialized cells are essential for many proliferative and differential functions of immune cells. Some cytokines may have functions that overlap with other cytokines. Nevertheless, individual cytokines act distinctively, sometimes synergistic (two or more cytokines acting together), and sometimes antagonistic (cytokines causing opposing activities) to each other [1].

Characteristic sets of cytokines are mainly produced by immune cells such as macrophages, B cells, T cells, and granulocytes, as well as several other cell types. As such, cytokines interact in great complexity with one another. These interactions represent a very sophisticated and versatile communication network, and are vital for the immune system to function effectively [2]. The physiological importance of cytokines is comparable to that of other signalling systems such as neurotransmitters and endocrine hormones.

It can be stated that the new (and growing) understanding of the biological mechanisms governing cytokine actions are an important contribution to medical knowledge [3]. The biochemistry and molecular biology of cytokine actions explain some well-known and sometimes also some of the more obscure clinical aspects of diseases. If manipulated correctly, cytokines might become an important weapon against diseases as infectious organisms grow increasingly resistant to antibiotics [4].

The race to the discovery of a gene or a drug has now become increasingly dependent on how quickly a scientist can scan through voluminous amount of information available online to construct the relevant picture (such as cytokine interaction pathways). Pathways are the logical format for modelling and presenting information in a manner that is familiar to biological researchers.

## 2 INTRODUCTION

The Cytokine Information System (CIS) provides the user with information on how a particular cytokine(s) react with other cytokines or other types of cells. It does this by visualizing it in the form of a graphical pathway. A simple pathway is as illustrated in Figure 1.

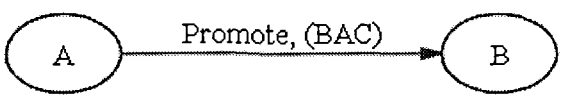


Figure 1: A Simple Graphical Pathway

*In Figure 1, Cytokine A is said to promote Cytokine B by* means of a BAC *catalyst.* The colour of the arrow represents the type of relationship between the two reacting cells as depicted in Table 1.

| Colour | Relationship |
|---|---|
| **Red** | **Cytokine – Cytokine** |
| **Green** | **Cytokine – Non-Cytokine** |
| **Blue** | **Cytokine – Unknown Cell** |

Table 1: Type of Relationship shown by the Arrow Colour

## 3 SYSTEM DESIGN

### 3.1 Design of Cytokine Information System

The CIS comprises of 2 main modules. The first module, named Natural Language Processing (NLP), is used to perform extraction of useful information from online sources such as web databases, journals and essentially any webpage. It is used to determine how certain cytokines react with one another.

The second module is a web application module, in the form of a Java Applet, to provide pathway visualization functionalities for cytokine-cytokine relationships. This is named as Cytokine Pathway Visualization (CPV) as shown in Figure 2.

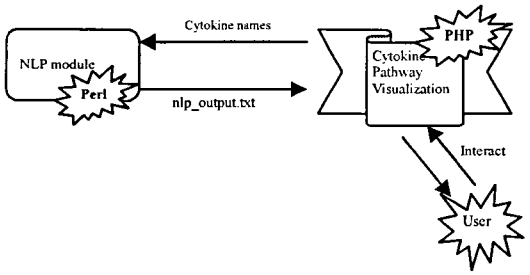


Figure 2 NLP module and CPV moduels

In additional to these main modules, a third module, named Cytokine Name Data Mining (CNDM), is used to perform data mining from online sources similar to that of the NLP module. It is used to determine the naming conventions and various aliases of existing cytokines. The system uses this module to obtain naming information of the cytokines specified by the user for processing. This feature is required because a common cytokine can have hundreds of aliases, and the best way to perform information extraction would be to use the most commonly used name.

## 3.2 Design of the Cytokine Information System

As the structured programming style was employed, the Data Flow Diagram (DFD) approach will be used in the following sub-sections to further illustrate the design structure of the CIS.

### 3.2.1 Level-0 Data Flow Diagram

The system takes in some inputs from the user via the main page of the website. The inputs include the number and names of cytokines to be used for processing. The user will be presented with a list of cytokines taken from the CNDM module's database. After the user has picked and confirmed the cytokines, the CIS will begin processing the inputs.

After processing, the results of the pathway visualization are displayed to the user via another webpage. The user may then perform additional functions such as saving the pathway visualization image, zooming, or further editing of the pathway. The Level-0 Data Flow Diagram (DFD) is as shown in Figure 3 and Data Dictionary in Table 2.

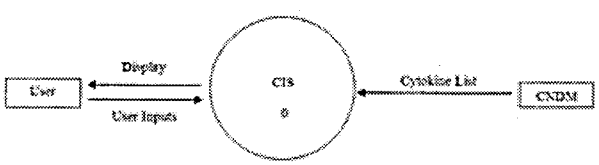


Figure 3: Level-0 Data Flow Diagram

| | | |
|---|---|---|
| **Display** | = | **Pathway Visualization Image + Additional Information from NLP** |
| **Intermediate File** | = | **Names Input by User + Additional Information from NLP + Relationships between Cytokines** |
| **User Inputs** | = | **Names Input by User + Number of Cytokines** |

Table 2: Data Dictionary for all DFDs

### 3.2.2 Level-1 Data Flow Diagram (CIS)

The Level-0 DFD is further broken down into a Level-1 DFD as shown in Figure 4. The CIS is refined into the two main modules – CPV and NLP.

The user inputs are first passed into the NLP module for processing. After its processing, it will output an intermediate file. This file contains all necessary information regarding the relationships, and is used by the CPV module to produce the resultant graph that is displayed directly to the user. The system interface of the CIS controls the flow of data and invocation of the two modules.

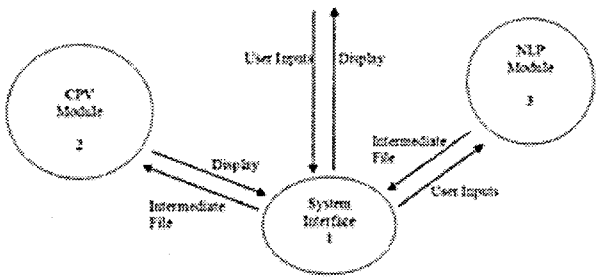


Figure 4: Level-1 Data Flow Diagram (CIS)

## 3.3 NLP module

Our NLP system consists of a simple web crawler and a NLP module to do the extracting information task. The NLP process has two main phases:

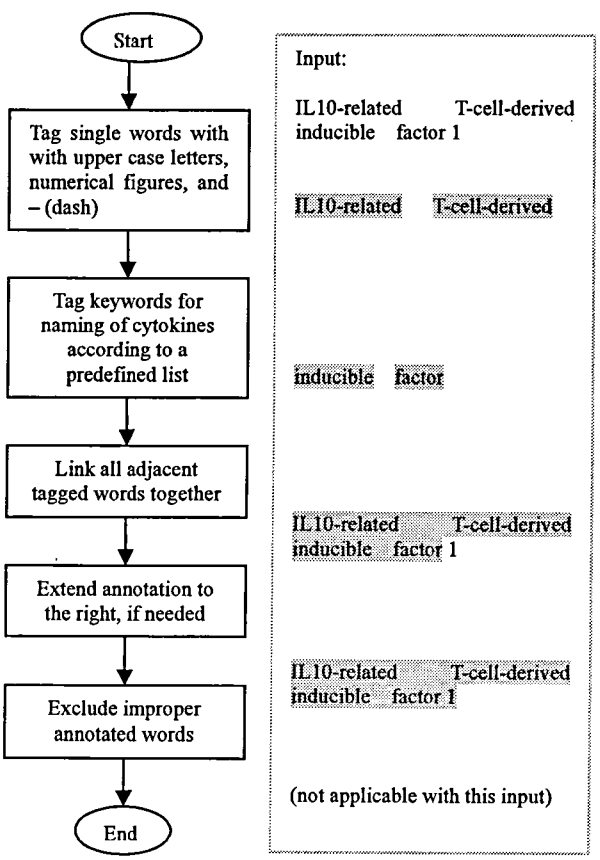


Figure 5: Tagging cytokine names against a set of matching rules

### 3.3.1 Tagging processing

This method uses a manually-constructed database consisting of 415 cytokine names/alternative names. To better cover the set of cytokine names not updated in the database, we exploit pattern matching templates [9] - [11] of cytokine names based on its systematical features of naming.

### 3.3.2 Parsing processing

We developed the context-free-grammar (CFG), as illustrated in Table 3 and 4, using Extended Backus-Naur Form. This set of grammar will cover a large set of interaction in biological articles.

| | |
|---|---|
| *Interaction* | Assignment \| Relationship |
| *Assignment* | *Noun_phrase* (NOT)? KEY *Noun_phrase* (AND KEY *Noun_phrase*)* |
| *Relationship* | KEY OF *Noun_phrase* (BY\|ON *Noun_phrase*)? ((KEY)+ Noun_phrase)? |
| *Noun_phrase* | ((NOT)? CYTOKINE)+ |

Table 3: Grammar in EBNF Form

| Non-terminal symbol | Description |
|---|---|
| Interaction | 'Start' symbol |
| Assignment | Assignment expression |
| Relationship | Relationship expression |
| *Noun_phrase* | Combination of cytokine names |
| ? | 0 or 1 occurrences of the specified quantity |
| + | more than 1 occurrence of the specified quantity |
| * | 0 or more occurrences of the specified quantity |

Table 4: Non-terminal symbol descriptions

Example 1:

The synthesis of **IL1** can be **induced** by other cytokines including **TNF-alpha , IFN-alpha , IFN-beta and IFN-gamma** and also by bacterial endotoxins , viruses, mitogens, and antigens.

Noun_phrase Key Noun_phrase

Extracted information:

| | |
|---|---|
| Sub: | TNF-alpha, IFN-alpha, IFN-beta, IFN-gamma |
| Action: | Induce |
| Obj: | IL1 |

The CFG could cover a large set of interaction of biological data. Although the grammar is comprehensive, but still could not detect all interactions, since human language is very complex. To detect the information, not covered by the CFG, we exploit a set of matching rules, which are kept updated to maintain the accuracy of NLP function.

Example 2:

In vitro **PGE2**, the synthesis of which is **increased** by **IL1**, and glucocorticoids **inhibit** the synthesis of **IL1**.

Noun Key Noun

AND Key Noun

Extracted information:

| | | |
|---|---|---|
| Sub: | IL1 | PGE2 |
| Action: | Increase | Inhibit |
| Obj: | PGE2 | IL1 |

We observe that to solve this false extraction, the word 'glucocorticoid(s)' need to be included into the list of non-cytokine noun. Then interaction glucocrticoids – inhibit – IL1 is correctly detected.

## 3.4 Cytokine Pathway Visualization Module

### 3.4.1 Level-2 Data Flow Diagram (CPV)

The Cytokine Pathway Visualization module is the main module that draws the graphical pathway. Its workflow is as illustrated as in Figure 6.

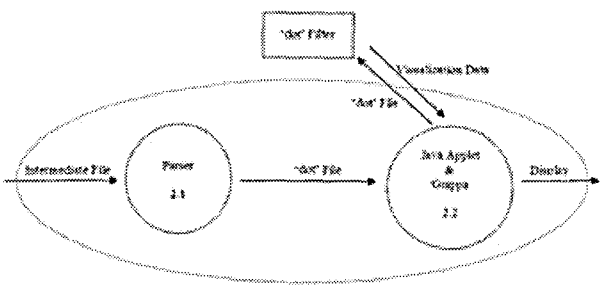

Figure 6: Level-2 Data Flow Diagram (CPV)

A parser, written in PHP, will first be run. It extracts the information from the intermediate file and outputs another file, which is of the 'dot' text format. This 'dot' file is required by Grappa, a Java graph package, to perform the necessary visualization functions.

When the 'dot' file has been written, a Java applet will be called. This Java applet is responsible for the GUI interfacing between the user and the Grappa package. The Java applet reads in the 'dot' file and then submits the contents to a 'dot' filter. The 'dot' filter is a program that can reside on the local server or any public web servers on the Internet. Data returned by the 'dot' filter is then used by the Grappa package for visualization of the graphical pathway.

### 3.4.2 PATHWAY VISUALIZATION

The Cytokine Pathway Visualization (CPV) module, written as a Java Applet, is responsible for the drawing and displaying of a graph. This graph is based on the output of the NLP module. Fundamentally, the graph depicts how various types of cytokines and non-cytokines are related to each other. Details of the relationships in terms of the type of relationship and catalysts involved are also shown. The user is allowed to modify the graph in a variety of ways and also to save a JPEG image of it as and when desired. The GUI of the CPV module is as shown in Figure 7.

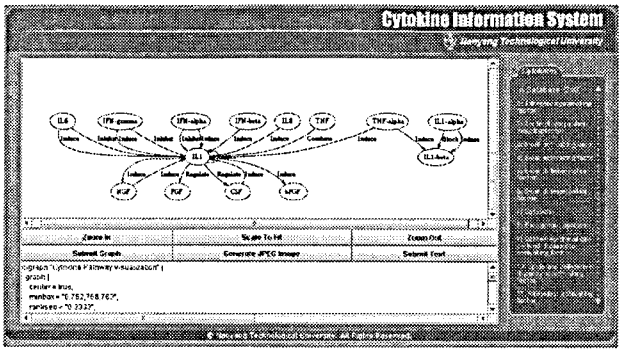


Figure 7: CPV Module invoked as part of the CIS webpage

### 3.5 Data Interface between CIS, CPV & NLP

Information is passed between the various modules and the CIS in the form of an intermediate file and a 'dot' file as shown in the DFDs in Section 3.2. These two files follow a specific format for the correct information passing and operation of the entire system. As long as they follow the correct format, the NLP and the CPV modules can be implemented in any way desired.

## 4. PROGRAMMING LANGUAGES

As there is no one single programming language that can cater to all requirements of this web application system, many different languages have been applied in order to produce accurate results and expected functionalities of the application. The system's several sub-modules have been written in different languages to perform tasks in the forte of the particular benefits of that programming language.

The entire application is built as a website, utilizing Apache Web Server, JavaScript, CSS, PHP, CGI Scripting, Java Applet, Grappa [5], and the DOT text format [6]. Several close interactions and information flow between these different software tools have been effectively demonstrated. It was also shown that when integrated well, these tools can provide for complex capabilities and solutions by each excelling in their own areas.

JavaScript and CSS were used mainly to layout the website and to create the forms that were used to allow the user to key in information. PHP was employed to provide dynamism and to control the information flow of the entire system. It was also used for file manipulation purposes. As Grappa is a Java package, a separate Java applet was written to make use of Grappa to perform the graphical pathway visualizations. Due to the applet being unsigned, CGI scripting was employed to work around some of its restrictions.

Grappa is a Java graph drawing package that simplifies the inclusion of graph display and manipulation capabilities within Java applications and applets. It has a good number of useful features built into it, but is also extensible. Grappa can be thought of as a port of a subset of GraphViz [7] to Java. It makes use of a DOT filter for drawing directed graphs, and is provided by GraphViz. The DOT filter works well on directed acyclic graphs (DAGs) and other graphs that can be drawn as hierarchies. It reads attributed graph files (written in the DOT text format) and writes drawings.

## 5. RESULTS

In the tests, the cytokine, Adherence-promoting factor, or better known as IL-1, is used. The IL-1 web page [8] from the Horst Ibelgaufts' Cytokines Online Pathfinder Encyclopaedia (COPE) was used for the natural language processing module. The result of the graphical visualization is as shown in Figure 6.

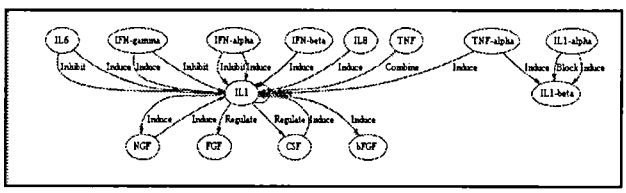


Figure 8: Graphical pathway visualization of the Adherence-promoting factor (IL-1) cytokine

As shown in Figure 8 and Table 5, the cytokines and keywords (type of action) have been extracted from the web page and drawn onto the pathway visualization. The red arrows depict a Cytokine-Cytokine relationship. Non-Cytokine-Cytokine relationships can be included but have been ignored in this simple demonstration presentation clarity. When referring to the web page from COPE, these relationships can indeed be inferred from the information given.

| Action | Cytokine(s) |
|---|---|
| Inhibited by | IL6, IFN-alpha, IFN-gamma |
| Induced by | IL6, IFN-alpha, IFN-beta, IFN-gamma, IL8, TNF-alpha, NGF, CSF, IL1 |
| Combined by | TNF |
| Induce | NGF, bFGF |
| Regulate | FGF, CSF |

Table 5: List of cytokines concerned with the Adherence-promoting factor (IL-1)

## 6. CONCLUSION

New (and growing) understanding of the biological mechanisms governing cytokine actions are an important contribution to medical knowledge. The biochemistry and molecular biology of cytokine actions explain some well-known and sometimes also some of the more obscure clinical aspects of diseases.

Furthermore, there are vast amounts of information on cytokine related reactions and relationships found on several online databases. In addition, even larger amounts of useful information exist amongst the published journals and papers on the Internet. This system is one that allows the user to screen through the entire web page and automatically visualize the key information into several graphical pathways. Its essence lies in its ability to quickly and accurately perform its task, thus saving time and effort on

the user's part.

This paper presented the Cytokine Information System and Pathway Visualization software as an effort to increase the proficiency of users in scientific discovery today.

## REFERENCES


[1] Mosmann T. R. (1991). Cytokines: Is there biological meaning? Current Opinion Immunol. 3, 311-314.

[2] Olsson I. (1993). The cytokine network. J. Intern. Med. 233, 103-105.

[3] Ben Z. Krentzman (2002). Cytokines. Online. Available: http://home.comcast.net/~bkrentzman 2005, March 10

[4] Palo Alto Medical Foundation (2003). Finding New Weapons Against Infection. Online. Available: http://www.pamf.org 2005, March 10

[5] AT&T Labs Research. Grappa - A Java Graph Package. Online. Available: http://www.research.att.com/~john/Grappa/ 2005, April 6.

[6] AT&T Labs Research. Grappa - The DOT Language. Online. Available: http://www.graphviz.org/cvs/doc/info/lang.html 2005, April 6.

[7] AT&T Labs Research. Graphviz. Online. Available: http://www.graphviz.org/ 2005, April 6.

[8] Horst Ibelgaufts' COPE: Cytokines Online Pathfinder Encyclopaedia. Online. Available: http://www.copewithcytokines.de/cope.cgi?5009 2005, April 6.

[9] K. Fukuda, T. Tsunoda, A. Tamura, T. Takagi, "Toward Information Extraction: identifying protein names from biological papers", 1998.

[10] Lorraine Tanabe and W. John Wilbur, "Tagging gene and protein names in biomedical text", Vol. 18 no. 8 2002, p1124-1132

[11] Gunnar Eriksson, Kristofer Franzen, Fredrik Olsson, "Exploiting Syntax when Detecting Protein Names in Text", Swedish Institute of Computer Science.