

# Oncosifter: A Customized Approach to Cancer Information

**Ketan Mane**

Laboratory for Applied Informatics Research  
SLIS, Indiana University  
Bloomington, IN 47401, USA  
+1 812 855 2849  
kmane@indiana.edu

**Sidharth Thakur**

Computer Science, Indiana University  
Bloomington, IN 47401, USA  
+1 812 855 3609  
sithakur@cs.indiana.edu

## ABSTRACT

This paper describes various interfaces used in the development of search engine called Oncosifter. Cancer related diagnosis & treatment, latest medical news and publications could be accessed using this system. Different interface modules provided for user interaction – keyword based, directory structure, hierarchical visualization interface, and personalized search are briefly discussed in the paper.

## Keywords

user-interface systems, cancer, oncosifter, cancer news, visualization user –interface, sifter

## INTRODUCTION

In the information age, a large volume of information is available in electronic format. This data spreads across various domains. Our system – Oncosifter, is adapted to focus on the cancer domain (“Onco” means cancer). A large amount of data on the research in labs and hospitals, news and diagnosis & treatment information is available. In this gamut of data, only a small fraction of information is of particular interest to the user. So it becomes increasingly important to have an automatic way information filtering. These techniques would eliminate the irrelevant information and present the user with all the relevant data based on their needs. Furthermore, data on the latest medical news, diagnosis & treatment and others are not available at the same website. This generates a need to develop a system that makes different kinds of data available. In addition to this, technical jargon associated with the domain makes it knowledgeable only to the medical community and the experts in that area.

Developing user-friendly interfaces, which map user’s needs to the information available, is one possible solution. However this approach demands increase in the user efficiency to search relevant data and standardize operations. In designing the Oncosifter system, an attempt has been made to address these concerns.

## SYSTEM DESIGN

Oncosifter is designed to retrieve the latest news and diagnosis & treatment information on cancer. It is targeted towards retrieving information from two cancer

information providing public websites: Medlineplus<sup>1</sup> and Cancer.gov<sup>2</sup>. Medlineplus is a source for all the latest news in the cancer research while Cancer.gov is used to acquire information about the different types of cancer, their detailed description, and diagnosis & treatment. The system is designed with focus on the non-technical community. Limited medical jargon is used with focus on the non-medical community. To ease the query submission process for less tech savvy individuals, user-friendly interface designs were adopted. Interfaces to provide generic or specific details were incorporated in designing the Oncosifter system. Consistency in the interaction style of different interfaces was maintained to shorten the learning curve to adapt to the system. Oncosifter was designed to keep interaction to a minimum and expedite the display of results.

The Oncosifter system is implemented using Perl-CGI. Different modules of interfaces used in the design of the Oncosifter search engine are as follows:

- Keyword Based Search Interface
- Directory Interface
- Hierarchical Visualization Interface
- Personalization Interface

Figure 1 shows an overview of the system. The following sections will provide information on layout, interaction styles, information processing and results display for each interface. The main interface (homepage) of the system is designed to incorporate keyword-based search. Other interfaces are linked from this interface.

## Keyword Based Search Interface:

The interface in this kind of search is comprised of a text box. Users specify their need in the form of a keyword. This is taken as a query and is matched with the keywords specified as metadata. If the query matches the metadata keyword, then the corresponding results are retrieved. However mapping information to a user’s interest may present a conceptualization problem [1], where concept

---

<sup>1</sup> <http://www.nlm.nih.gov/medlineplus/>

<sup>2</sup> <http://www.cancer.gov>

terms used to represent available information may differ from user query term [2].

In Oncosifter, a modified approach has been adopted for back-end query processing. Typically in the cancer domain, one keyword can be associated with a collection of a cancer group. For example: if the user queries for “BONE CANCER”, there exists two types of cancer in that group: Ewing’s family of tumors and Osteosarcoma/Malignant Fibrous Histocytoma of Bone. Thus it becomes essential to provide the user with details of both cancer types. In our approach, the query term is compared with the combination of keyword and cancer term URL addendums. If a match is found between the query and keyword, the corresponding terms are retrieved. These terms are used to retrieve information from above mentioned websites. For bone cancer, we have the corresponding match:

BONE CANCER#Bone@ewings@osteosarcoma

Wherein, BONE CANCER is a keyword delimited from the rest of the data. The terms after the delimiter “@” are URL addendum terms. The term “Bone” is used for group labeling purpose.

After retrieval, information filters are used to parse out relevant information. The results for each term are concatenated. This final data composed of all the relevant information is presented across to the user in customized format. Figure 2 shows the interface layout.

This kind of approach supports the retrieval of results based on medical vocabulary, body location, body-systems and commonly used terms. These terms would serve as matching keywords in the data structure.

#### **Directory Interface:**

This interface is mapped with the “Directory” word on the main page of the system. It provides an overview of the different types of cancer. For simplified search, the cancer types are categorized into three main sections: by body location/systems, common cancers and childhood cancers. Wherever possible common vocabulary terms are used. Within each term is embedded a URL that is used for dynamic retrieval of results. For example: In “BONE CANCER”, we have the following URL:

<http://oncosifter.indiana.edu/cgi/directory.cgi?Bone@ewings@osteosarcoma>

The CGI script – directory.cgi is used to process URL information. Apart from clicking on the term of interest, a similar approach of retrieval, filtering and adding together all the results is adopted. However the results are displayed to the user in a separate browser window. Additional information on the order and count of the cancer results is available at the top of the page. Figure 3, shows the directory interface design.

The concept of different results popping up in different windows was adopted for efficient user interaction. The results retrieved include information on the different

stages of cancer and diagnosis & treatment. Multiple browser display would aid in comparing the cure treatment options available. Also consistency in information layout is maintained to reduce the cognitive load on the user [3].

#### **Hierarchical Visualization Interface:**

Graphical visualization(s) of the data set helps revealing underlying structure in the data, which is difficult to achieve by direct analysis of the data. Furthermore, they are helpful in displaying structural relationships in the data [3]. Effective visualizations strive to comply with the information seeking mantra – Overview, Zoom and filter, and Details on Demand [5].

The cancer categories represent a hierarchical tree data structure. No interrelation exists within different subtrees. Higher levels of data are parent and sub-levels belonging to the same tree branch are considered as children. This concept is followed throughout the data structure. Hyperbolic tree visualization is one of the common layouts used for such a kind of data structure. In addition, this kind of visualization helps to maintain the user’s location in the information space. This feature can be explored by clicking on the word “Visualization” from the main page.

In the following data set, the body location/systems, common cancers and childhood cancers act as parent nodes for their individual categories. In one level below, the group terms serve as parents to the different types of cancers within the group. Visualization also has explicit color-coded nodes to provide navigational cues. A color-fading feature is used for visual identification of the node levels. Child nodes are more lightly colored than their parent nodes. An overview snapshot of the hyperbolic tree layout is shown in Figure 4.

The final level of node comprises of different cancer types. These nodes are click able URLs to the corresponding cancer information. The page is parsed for relevant information using CGI script and results are always presented in a new browser window.

This visualization is made portable by implementing it as a Java-based applet and it assumes that the user’s browser is Java applet compliant.

#### **Personalization Search Interface:**

User-profiles can be created using this interface. In Oncosifter, “user-profile” means include information that is of interest to the user. Individual profiles can be created by filling in a username and desired password information. A typical error check is performed and relevant feedback on the missing information is given to the user. Once the sign-in and profile is created, individual user’s can access their profiles through typical signing process.

In Oncosifter, choosing the cancer terms from the intermediate interface can create profiles. It also presents

a rating scale of ten to acquire information on user interest in certain topics. Figure 5 shows the layout of this interface. A descending ordered list is generated based on user topic preferences. Most interesting term results are displayed at the top. The user is also given the option to edit their profile. Additional terms of interest can be added and the irrelevant data can be deleted. Categorization of cancer news obtained from the website is done using these terms. The news titles are embedded with the URLs to the article. By clicking on these links, information retrieval and filtering is done on the article.

Along with the article a rating scale is provided. Based on this feedback, automatic changes [6] are reflected in the user's profile.

**CONCLUSION**

Oncosifter provides access to different types of information at the same location. Limited use of medical jargon makes it favorable for non-medical experts. Consistency and common interaction style is maintained throughout the system. It has been successful in achieving its goal of keeping the interaction to bare minimum and provide instant data access.

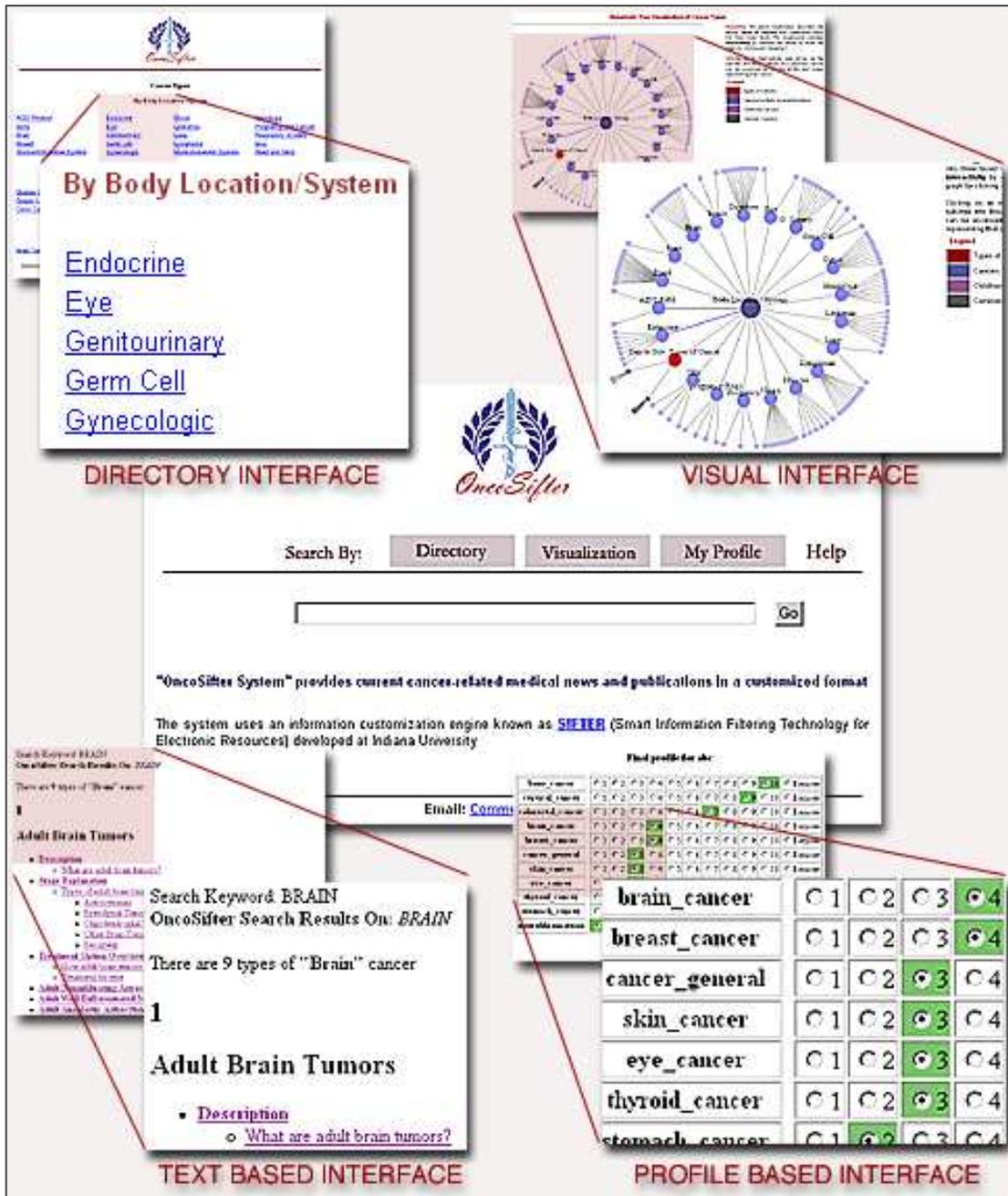


Figure 1: Overview of the Oncosifter search system displaying the various search interfaces



Figure 2: Homepage and Keyword based Search Interface



Figure 3: Directory Search Interface

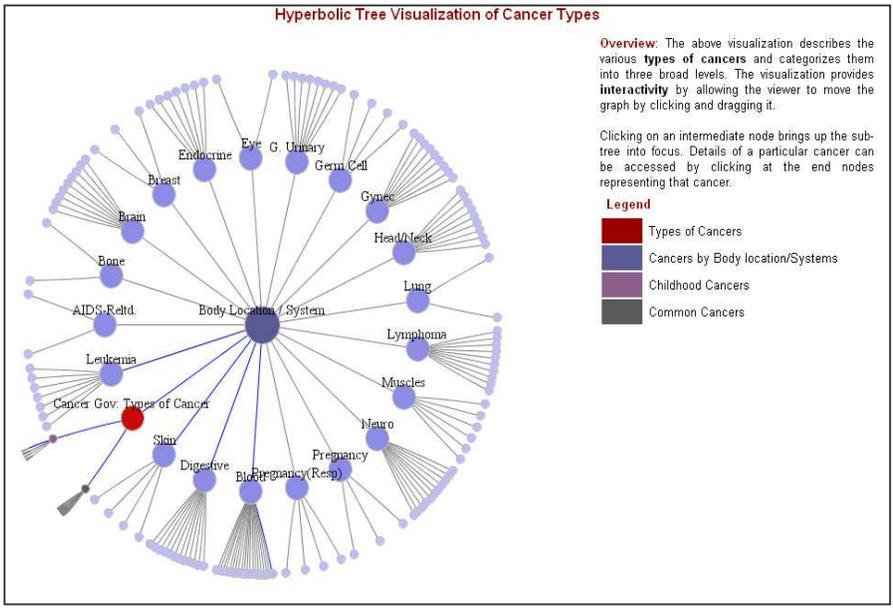


Figure 4: Visualization Search Interface

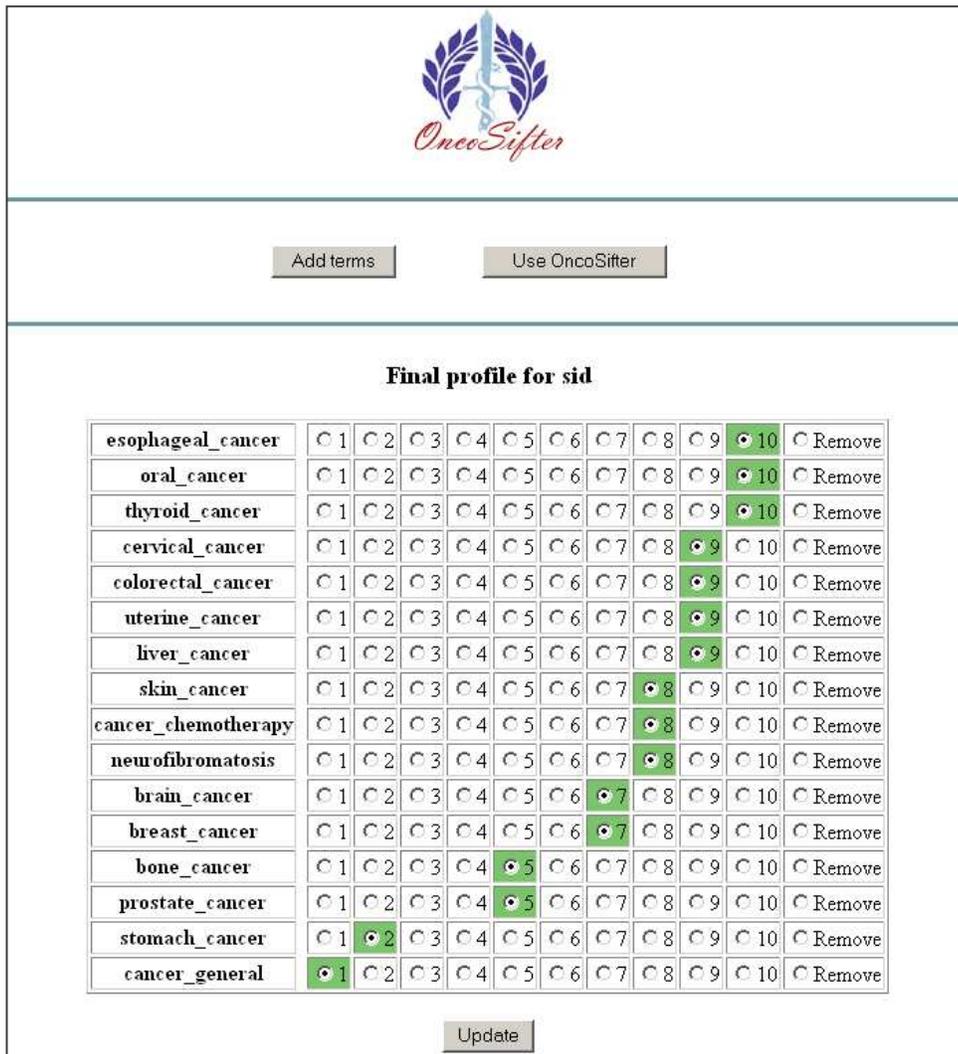


Figure 5: Personalization Search Interface

#### ACKNOWLEDGEMENT

We would like to thank Dr Javed Mostafa and Raghveer Mukhamalla for providing valuable insight during the design process of the OncoSifter.

#### REFERENCES

1. Furnas, G.W. Landauer, T.K, Gomez L.M and Susan Dumais, S.T. (1987), The vocabulary problem in human system communication. *Commun. ACM*, 30(11): 964 – 971
2. Gaines, B. R. and Shaw, M.L.G, (1989), Comparing the conceptual system of experts, In *Eleventh International Conference on Artificial Intelligence*, 633 – 638
3. Robertson, G. G., Card, S. K., Mackinlay, J. D., (1993). Information Visualization using 3D Interactive Animations, *Commun. ACM*, 36(4), 57 – 71
4. Foltz, P. W., Dumais, S. T., (1992), Personalized Information Delivery: An Analysis of Information Filtering Methods, *Commun. ACM*, 35(12), 51 – 60
5. Shneiderman, B. (1997). Human factors of interactive software. In *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Addison-Wesley, 1-37.
6. J. M. Mostafa, S. Mukhopadhyay, W. Lam and M. Palakal, (1997), A Multilevel Approach to Intelligent Information Filtering: Model, System and Evaluation, *ACM Transaction of Information System*, 15(4).