International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 3, May-June 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Dynamic Categorization of Biomedical Search Results

[1]Ulasa Venkateswarao, [2]Mrs .M.Anuradha
[1](Student Dept. of Master of Computer Applications, Amrita Sai Institute of Science and Technology, Paritala, Andhra Pradesh, 521180, India.
Email: venkateswaraoulasa1@gmail.com)
[2](Asst.Prof, Dept. of Computer Science & Engineering, Amrita Sai Institute of Science and Technology, Paritala, Andhra Pradesh, 521180, India.
Email: anuradhanallagonda91@gmail.com)

----------**********----------

Abstract:

Because the number of biomedical publications is rising quickly, it becomes difficult to efficiently obtain the needed information for major biology research by using repositories such as MEDLINE, Entrez Gene and OMIM. Using traditional PubMed–like search tools often makes users deal with too much information at once. As a result, users have to keep tweaking their searches, possibly overlooking important studies and it can get tiring mentally. BioNav is now used to group search results by using the detailed Medical Subject Headings (MeSH) ontology as a tree structure. Browsing in this way is made easy by each node in the tree being a MeSH concept, helping users explore generic to particular subjects, as can be done when navigating product categories on e-commerce sites. When using BioNav, citations are matched to relevant concepts through MeSH tags and through detecting terms in the text and the nodes are arranged by how often the citations appear and their relevance. Because of this strategy, researchers find it is easier to locate studies about prothymosin in different fields of biomedical research.

Keywords — *Biomedical literature, information overload, MEDLINE, PubMed, MeSH ontology, BioNav, categorization, navigation tree, prothymosin, literature retrieval, biomedical informatics.*

----------**********----------

1. INTRODUCTION

Because of the fast-paced growth of biomedical literature, MEDLINE (which supports PubMed) currently stores over 18 million citations and adds approximately half a million new entries every year. The same growth is happening in other important gene databases which means researchers now need to handle both new chances and new difficulties as the area grows. Now it is very important to search and analyse this huge corpus of data for planning and interpreting important biological projects.

Normally, biomedical researchers—comprising biologists, chemists and health scientists—use keyword searches on PubMed to find important papers. Still, in the exploratory stages, when precise keywords have not been set, users will normally use general searches that bring back a lot of results. Using targeted keywords might find less, but it may also miss some key articles in the literature. To give an example, searching for "cancer" in PubMed gives over two million citations, whereas finding "PR thymosin" shows only a few hundred, so deciding what is important can become very difficult.

Such a situation underlines the frequent issue called information overload and a number of answers have been suggested to address it. Most of these approaches fit into two groups: ranking and categorization and some use elements of both. BioNav relies mainly on categorization, an approach that is suitable for biomedical research as many big, carefully built concept hierarchies exist such as the Medical Subject Headings (MeSH) ontology.

Using ranking based on how lightweight categories are organised, BioNav improves the user experience and the quality of search results. All search results are displayed in a special structure called the navigation tree organised by subject areas. People can look at different areas of computer science, starting with top-level topics and narrowing in on what they want.

Every MEDLINE citation is given MeSH terms by being annotated by editors and by including textual hints. BioNav uses both to link mentions of citations with the matching MeSH terms. The user can browse the root concepts, sorted by how many times they are cited and once any are expanded, the root shows the child concepts, sorting them by importance or frequency, until the topic and related papers are discovered.

Amazon and eBay are online shops that work in this way, so this model is inspired by their approach. Through this process, researchers can look into various topics, for example those about prothymosin, by digging into citations organised by subject branch.

2. SYSTEM ANALYSIS

2.1 Identification of Need

The search engine PubMed which uses keywords for biomedical searches, can create difficulties for those who use it for exploratory queries. Usually, when a user makes a basic query, they get numerous outcomes, so to pare down the list the user has to fine-tune their search using more focused keywords. Because of trying to narrow the search too much, the process might not find all the appropriate citations.

2.2 Limitations of the Existing System

Current databases in biomedicine such as PubMed, are overwhelmed with large amounts of information. By using targeted queries you can still find thousands of papers and only a few will be truly useful to you. Lacking convenient ways to explore results in more missed important papers and a bigger need for users to handle things manually.

2.3 Proposed System: BioNav

Organising the query results with BioNav makes use of the Medical Subject Headings (MeSH) ontology to improve the understanding of the main concepts. Instead of relying on tuple-based structures and trying to minimise navigation expenses like prior systems, BioNav uses the existing MeSH hierarchy which works best in biomedicine.

Biomaterials Navigator organises query results into a tree and each node in the tree stands for a MeSH concept. Users can jump from general topics to individual subtopics which improves how easy and organised the literature research is. BioNav is different from static navigation systems because it offers changing, user-centred navigation which ensures greater adaptability and better user experience.

2.4 System Modules

- **Query Search Process / Biomedical Search Interface**: Users can type in their search keyword(s) in the same way they would in a normal PubMed search. Citations are grouped using the MeSH hierarchy which may be done manually or by recognising terms in the texts.
- **Dynamic Navigation Tree**: Organises the outcomes of a search so the results are grouped by MeSH concepts. The root shows major concepts organised by how frequently they are cited and users are able to open nodes to see all the ranked

concepts below them, not only the immediate ones. This focused increase of topics enables easy exploration and helps by using popular e-commerce navigational ideas.

- **Hierarchy Navigation Web Interface**: Puts query results into a well-structured hierarchy, so users can focus only on what is important to them and skip unimportant topics. Each node is named clearly so it is easy to browse the MeSH groups by using their familiar labels..
- **Query Workload Online Operation**: BioNav uses the Entrez Programming Utilities (eUtils) to collect relevant citation information from MEDLINE after a query is received. After that, it gathers all the citations and links them with the related MeSH concepts, using the tree identifiers to encode how each is organised. Every user query goes through this process to provide the latest organisation of pages.

2.5 DYNAMIC NAVIGATION TREE

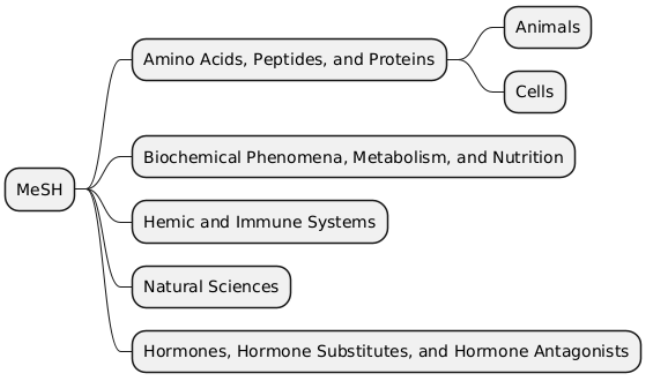


Fig: Navigation Tree

2. REQUIREMENT SPECIFICATIONS

2.1 Software Requirements

Component	Specification
Front End	J2EE (HTML, Java, JSP, Servlet)
Application Server	Tomcat 5.0 / 6.X
Scripts	JavaScript
Server-side Script	Java Server Pages
Build Tool	Ant
Database	MS-Access
Database Connectivity	JDBC
IDE	NetBeans 6.0.1
Operating System	Windows Family

Table 1: Software Specifications

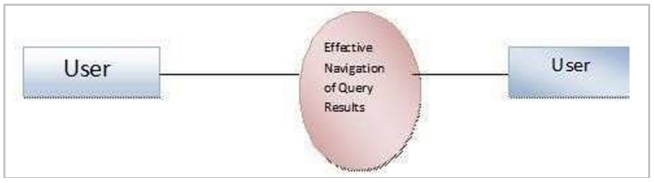
2.2Hardware Requirements

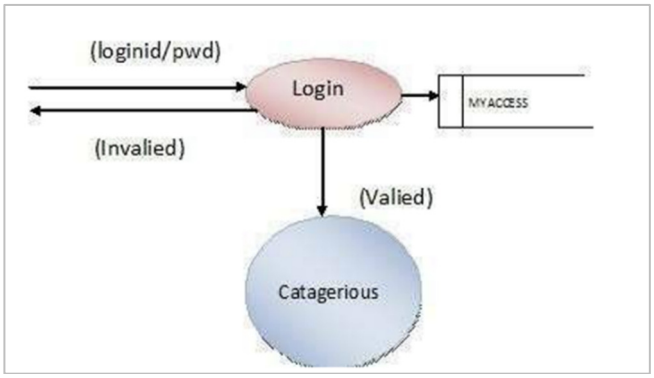
Component	Specification
Processor	Pentium IV 2.6 GHz
RAM	256 MB
Hard Disk	20 GB

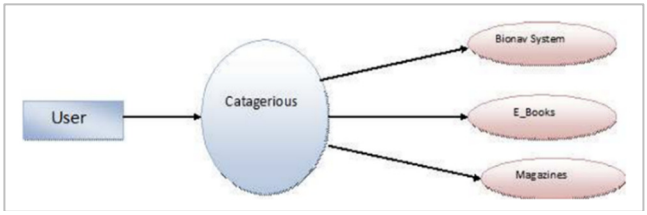
Table 2: Hardware Specifications

3. SYSTEM DESIGN

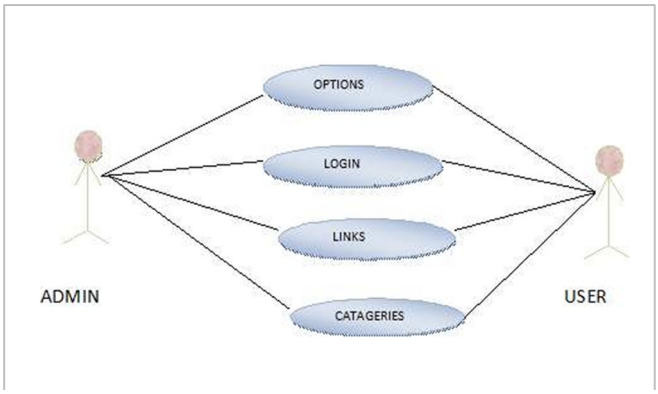
3.1 Context level diagram



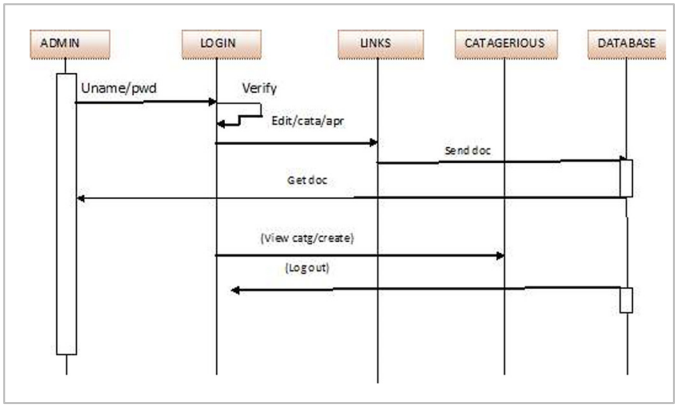




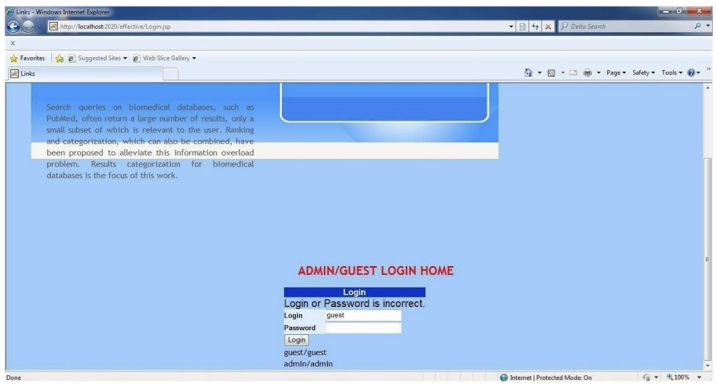
3.2 USECASE DIAGRAM



3.3 SEQUENCE DIAGRM



4. VALIDATION CHECKS



5. TEST CASES

Test Case ID	Input Description	Expected Result	Pass/Fail
NQR_TC_01	Login ID blank, password blank	Login ID and password fields are mandatory	Error
NQR_TC_02	Valid ID and blank password	Password is mandatory	Error
NQR_TC_03	Password and blank ID	Login ID is mandatory	Error
NQR_TC_04	ID < 4, valid password	ID should contain more than 4 characters	Error
NQR_TC_05	Valid ID and password < 6	Password should contain a minimum of 6 characters	Error
NQR_TC_06	ID with $, %, #, password	ID should not contain special characters	Error
NQR_TC_07	ID with spaces and password	Spaces are not allowed in ID	Error
NQR_TC_08	Invalid ID, invalid password	Invalid ID and Password	Error
NQR_TC_09	Valid ID, valid password	Valid ID and valid password are given	Login Successful

6. REPORTS

Fig: Home Page

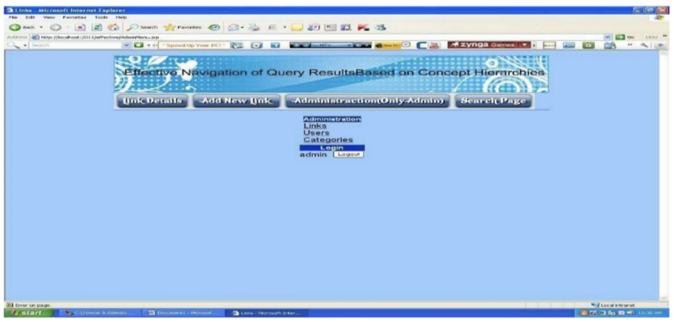


Fig: Admin Home Page

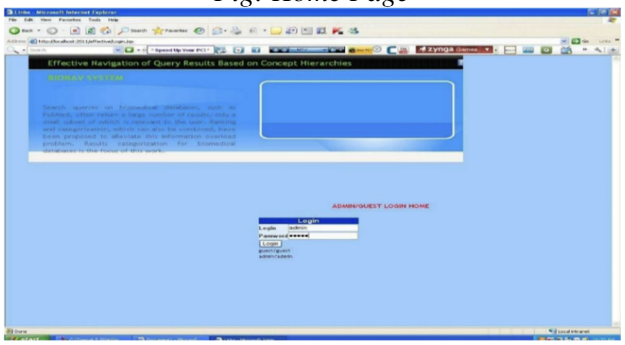


Fig: Admin Login Page

Fig: Add New Catageory

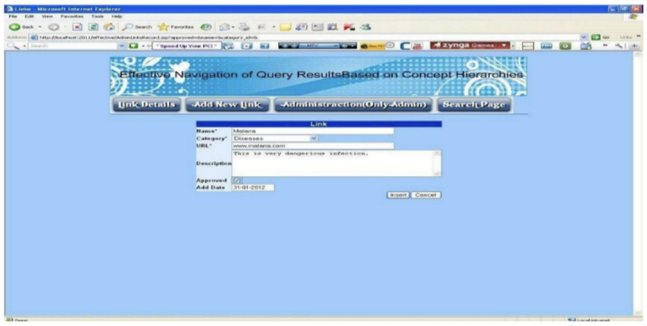


Fig: View Users

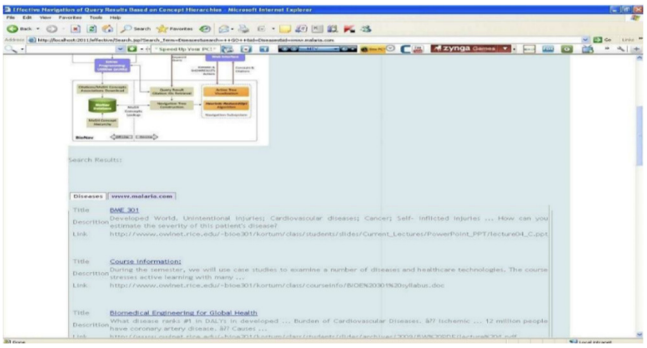


Fig: Search page

7. CONCLUSION

In short, BioNav deals with the challenge of too much biomedical information by organising it with a layered structure of MeSH terms. Because BioNav organises and philtres results by relevance, users can navigate the data with less mental effort. We set up a proper structure that included plans for navigation and how costs are handled and we showed that determining the best expansion of the route tree is a hard problem. Both a fast algorithm for short trees and a quick heuristic for regular usage were introduced to deal with this. Tests conducted on the approach strongly showed that it is superior to categorization only based on published details. Using the BioNav system architecture in practise strengthens its chances of success in real biomedical information retrieval applications.

ACKNOWLEDGEMENT

I am thankful to the Management of Amrita Sai Institute of Science and Technology for giving me an opportunity to work with his project.

I would like to the thank **Dr. M. Sasidhar**, Principal, Amrita Sai institute of science and technology, for his constant encouragement and support during the progress of this work.

I am deeply grateful to **Dr. P. Chiranjeevi**, Professor and Head of the Department, for his valuable guidance and consistent support during the course of the project.

A special note of thanks to my internal guide, **Mrs .M.Anuradha (M.Tech)**, for her exceptional guidance, constant motivation, and continuous encouragement, which played a crucial role in the successful completion of this project.

ULASA VENKATESWARAO

REFFERENCES

[1] J. S. Agrawal, S. Chaudhuri, G. Das, and A. Gionis, "Automated ranking of database query results," in *Proc. 1st Biennial Conf. Innovative Data Systems Research (CIDR)*, Asilomar, CA, USA, 2003.

[2] K. Chakrabarti, S. Chaudhuri, and S. W. Hwang, "Automatic categorization of query results," in *Proc. ACM SIGMOD Int. Conf. Management of Data (SIGMOD)*, Paris, France, 2004, pp. 755–766.

[3] Z. Chen and T. Li, "Addressing diverse user preferences in SQL query-result navigation," in *Proc. ACM SIGMOD Int. Conf. Management of Data (SIGMOD)*, Beijing, China, 2007, pp. 641–652.

[4] L. Comtet, *Advanced Combinatorics: The Art of Finite and Infinite Expansions*, rev. enl. ed. Dordrecht, Netherlands: Reidel, 1974, pp. 176–177.

[5] R. Delfs, A. Doms, A. Kozlenkov, and M. Schroeder, "GoPubMed: Ontology-based literature search applied to Gene Ontology and PubMed," in *Proc. German Conf. Bioinformatics (GCB)*, 2004, pp. 169–178.

[6] D. Demner-Fushman and J. Lin, "Answer extraction, semantic clustering, and extractive summarization for clinical question answering," in *Proc. COLING-ACL*, Sydney, Australia, 2006, pp. 841–848.

[7] National Center for Biotechnology Information, "Entrez Programming Utilities," [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query/static/eutils_help.html
[8] U. Feige, D. Peleg, and G. Kortsarz, "The dense k-subgraph problem," *Algorithmica*, vol. 29, no. 3, pp. 410–421, 2001.
[9] V. Hristidis and Y. Papakonstantinou, "DISCOVER: Keyword search in relational databases," in *Proc. 28th Int. Conf. Very Large Data Bases (VLDB)*, Hong Kong, China, 2002.
[10] R. Hoffman and A. Valencia, "A gene network for navigating the literature," *Nat. Genet.*, vol. 36, no. 7, p. 664, 2004.
[11] Humboldt-Universität zu Berlin, "Ali Baba: PubMed as a graph," [Online]. Available: http://alibaba.informatik.hu-berlin.de/
[12] Information Hyperlinked over Proteins (iHOP), "iHOP - Information Hyperlinked over Proteins," [Online]. Available: http://www.ihop-net.org/UniPub/iHOP/
[13] A. Kashyap, V. Hristidis, M. Petropoulos, and S. Tavoulari, "BioNav: Effective navigation on query results of biomedical databases," in *Proc. IEEE Int. Conf. Data Eng. (ICDE)*, Shanghai, China, 2009. [Online]. Available: http://www.cs.fiu.edu/~vagelis/publications/BioNavICDE09.pdf
[14] S. Kundu and J. Misra, "A linear tree partitioning algorithm," *SIAM J. Comput.*, vol. 6, no. 1, pp. 151–154, 1977.
[15] W. Lee, L. Raschid, H. Sayyadi, and P. Srinivasan, "Exploiting ontology structure and patterns of annotation to mine significant associations between pairs of controlled vocabulary terms," in *Proc. Data Integration in the Life Sciences (DILS)*, 2008, pp. 44–60.

