# PifMed: A System for Categorical Exploration of MEDLINE Articles using the MeSH Taxonomy

Pif Edwards
Department of Computer Science
University of Toronto
pif@cs.toronto.edu

Vlado Kešelj
Faculty of Computer Science
Dalhousie University
vlado@cs.dal.ca

## ABSTRACT

The subjective nature of human thought, preference, and sense of relevance will continue to thwart ranked lists and predictive agents, regardless of the quality and quantity of data. We can side-step these human obstacles by viewing search as collaborative process between humans and computers and allowing humans to use their subjective judgement in navigating an objectively organized tree of results. With the application of knowledge-based methods we can categorize the results of search, but then also use these hierarchical categories as the basis for a browsing tool.

We present theoretical and experimental evaluation of an IR system for retrieval of medical articles. The PifMed system presents MEDLINE results to the user, categorized by the MeSH terminology and hierarchically browsable. Practical evaluation conducted by within-group controlled experiments of a prototype system show a mean speed increase of 26 seconds and a strong user preference when directly compared to PubMed. Theoretical evaluation argues a move from lists to trees, is a move from a browsing structure to a navigation structure, and likens the shift to parallel the shift from list search to tree search, where the users' subjectivity is the comparator.

## 1. INTRODUCTION

The problems of linguistic ambiguity persist and continue to thwart AI research. Even if these problems could be solved or an effective compromise could be found, the human subjective sense of relevancy is an obstacle with no solution. Even if computers could read minds, automatic agents would still be stymied by the common condition every searcher has faced "I don't know what I'm looking for, but I'll probably know it when I see it." The fact is humans are unpredictable and searches cannot — nor will ever — be completed without cognitive effort by the user. Search must be seen as a collaborative task between human and computer, a dynamic process, not a packaged product.

The information gathering process has an artificial boundary between the query process and the browsing process, that is, browsing often leads to query refinement. Conventional search reinforces those boundaries. If we can blur the artificial boundary between query and browsing, by making all query refinement an implicit part of the browsing process we can deepen the collaboration between human and computer in IR. When we blur the semantic divide between user and computer (as the Semantic Web does) we are given an opportunity to make query refinement implicit and we enter a new phase of human computer interaction. When we organize information in the way humans think, humans can think their way through the information.

The Semantic Web as foreseen [1] is a curated web, curation and categorization are essential to traditional libraries and many systems of organization have been developed by libraries for this purpose, thus we offer a glimpse into 'Web 3.0' by taking a step back into library science. As the web starts to behave more like a library, we thought look at libraries for how to progress the Web. In this paper we show how curation and knowledge-based methods can be used in interface design to direct users to relevant items, faster and in a way they prefer over conventional lists.

One problem for categorization is the difficulty and expense of obtaining high-quality, gold-standard data. Within the medical domain we a have a unique opportunity to take advantage of a large, freely available, expertly curated, hierarchically categorized, linguistically sophisticated document corpus: MEDLINE. This corpus is categorized into a mature, domain-specific, annually-updated taxonomy called MeSH by domain experts at the U.S. National Library of Medicine (NLM). This database holds over 19 million items, each indexed with dozens of terms from a meticulously-created hierarchical system of over 26,000 descriptors. It is of the highest quality, there are millions items and it is freely available, thus with the usual constraints to categorization removed, practical analysis (through strong experimental methods) can be used to directly compare the conventional search interface to a search interface only possible with a human-curated collection.

## 2. RELATED WORK

Categorized hierarchies are nothing new to the Web. There several examples of browsable category hierarchies on the WWW: Wikipedia [13], Yahoo! Answers [14], Open Directory Project [9] which has been implemented into Google Directory (www.google.com/dirhp), and U.S. Patent Database [11] to name a few. There are many hierarchical classification systems used by library systems, for example the Dewey Decimal System [12], Library of Congress Classification, ACM Computing Classification System, and the Medical Subject Headings (MeSH) used by the U.S. National Library of Medicine are some of the important taxonomies. If the reader has an interest in related systems, I would direct them to an excellent survey paper on the subject of web clustering engines [2].

## 3. CATEGORICAL EXPLORATION

Ranked lists with large result sets leave lower ranked results disregarded by users. That is, when 100s of items are returned for a given query, the results further down the list are practically unreachable to the user. With a long list the user has little choice for exploration: *grind through* every result; or *skip down* the list until they find a relevant item; or give up, *refine the query* and start from scratch. We offer a solution that makes large result sets practical for human use: **categorical exploration.**

We examine a hierarchical category tree model in place of a ranked list model. In this model, the user is presented the results in a interactive browsable tree where the *nodes are category names* and the *leaves are articles*. The parent-child relationships between nodes are the hypo/hypernymic relationships pre-defined in a hierarchical category system (such as the MeSH Taxonomy). The user browses the system in the same way they would a file hierarchy in Windows, Linux or Mac OS.

## 3.1 From Browsing to Navigation Structure

By shifting from a ranked list to a browsable tree, we move from a **browsing structure**, to a **navigation structure**. Navigation requires a 'geography': stars, street signs, landmarks or maps; a way for persistent features to guide the user. A list algorithm creates a substantively different, wholly disposable ordering for each different query phrase, however, a hierarchy persists despite the query. Thus the hierarchical taxonomy acts as a persistent 'geography' to be comprehended and remembered. Each search session shows paths through the tree which can be remembered to aid future travels. Here is a tool with a knowable, predictable, sensible underpinning. Only with this system can we navigate a result set – the hierarchy (e.g. MeSH) provides the map. Lists, on the other hand, are constrained to one path, browsed and skimmed in only one direction, down.

In a list each item must be examined in linear order, whereas (like any tree) the order in this tree is not linear. Thus the search path is not *dictated by the structure* but *determined by the user*. In fact, like any tree, each article, by comparison, is very close to 'the top of the list'; each article (in MeSH) is less than 10 steps from the root. And since each article is in multiple categories, there is more than one path to each article. In a list any duplication of articles would be frowned on as needless redundancy, here we can see it as beneficial. Unlike a list, this non-linear structure makes no judgement on 'the best order', but allows each user to find their own order of visitation, and only with this system are different paths to the same result possible.

## 3.2 Tree Structure Versus List Structure

Instead of a **list** browsing structure, we use a **tree** browsing structure. A parallel can be seen between computer search and human search. Think of the list data structure vs the tree data structure. Each item in a list must be viewed sequentially, this *linear* order of visitation is dictated by the structure. In a tree, the order of visitation is *non-linear*, a series of choices determine the path. For tree search we have a well-defined objective **ordering** (alphabetical), objective **comparator** (greater/less than), and an objective **end-state** (exact match).

If we see the human as part of the algorithm, and the users' subjective judgment as the comparator, then we can see the category tree structure as enabling tree search for information retrieval. At each node in the tree (category), the user chooses to visit the node (i.e. see its children) or chooses to move to a sibling, just as the comparator in tree search makes a choice at each node. Each leaf in the tree (article) is a possible match for the end state, but this matching is complex and subjective to each user. Since only the user can judge relevancy at each node and leaf, it is beneficial to let the user into the problem-solving mechanism, into the algorithm. When the human and computer co-operate this way, strengths of each benefit the task. If the human is taken as part of the algorithm, we have an arguably objective ordering (hypo/hypernymic hierarchical categorization), a subjective comparator (more/less specifically relevant judgement of category) and an subjective end-state (satisfaction of information need).

## 3.3 Subjectivity as Guide Instead of Obstacle

Objectivity still has a major role in this system, but rather than judging relevancy in relation to the query, it is in judging relevancy of hypo/hypernymic relationships (known as ISA relationships) in the categorization hierarchy. Determining if two categories have a hyponymic or hypernymic relationship is much less subjective than if Order X is the best ordering of results given Query Phrase Y. Consensus is possible for the majority of categories and near consensus for all but a few. This can be supported by the popularity and successes of WordNet and intensive work on ontologies.

## 3.4 Implicit Query Refinement

Each category over which the user browses in an opportunity to focus the formulation of the query. By blurring the lines of distinction between query formulation and browsing we can make the search process more responsive to human thought. There is an opportunity in a hierarchical categorization tree for the user to merge query refinement and article selection into the *same action*, browsing.

Even experienced users of search engines have used a query phrase which was 'the best they could do at the moment' rather than 'the best that could be done', that is, used a query that quickly came to mind — just to get started — because the right words were elusive. It is these queries in particular that would benefit from query refinement. An important point is that new query phrases need not be entered, the system needs no feedback from the user to produce a new set for the user to browse. The refinement is an *implicit* function of the browsing process.

## 3.5 Relationships between Results

There is a dimension of query search results which is ignored by ranked lists. A ranked list, ranks solely on the dimension of relevancy to the query, that is, only the relationship to the query informs the ordering. There is another important dimension which can be used as the basis of organization: *the relationship of each result to each other result.* All results are related to the query, but in the category tree they are presented to the user in how they are related to each other, in categories. These categories are then related to each other to form a tree. A ranking system assumes independence in this dimension for simplicity sake, focusing instead on guessing the subjective needs of user based on the query.

Likewise, two users may use identical queries for completely different information needs. The query is a moving target, yet when we rank, we rank based on query. That means every identical query produces an identically ranked list. The subjective nature of human perception causes problems for these objective computational models; no matter how good they are, no one objectively ordered list will likely satisfy the subjective information needs of all users. Furthermore, when we rank based on query alone, our ranking is only as good as the query. Many users create poor queries because they have only a vague idea of what they are looking for, thus the query is often a poor match for the information need. All the assumptions of ranking systems fail when this is the case. Since a hierarchy does not rank the results, it makes no assumptions of this kind, and the user can first browse categories (not articles) for relevancy, and upon finding one, can then continue searching from a 'more relevant footing'.

This method is particularly beneficial to searches which require high recall. Exploration being categorical, the user can use these well-defined boundaries to isolate ideas, methods and perspectives, then sample from semantically different categories as an exploration strategy to reveal the breadth of a result set without re-querying the system.

## 4. PIFMED SYSTEM

PifMed organizes MEDLINE search results into a browsable tree according to the MeSH categories which have been assigned to each item. A screenshot of PifMed is shown in **Figure 1**. PubMed is the internet interface that searches MEDLINE and presents the results in a conventional ranked list. In this section we describe some details of these MEDLINE, PubMed and MeSH as they pertain to this implementation and this experiment.

*Index Medicus*, created in 1879, was a comprehensive index of medical journal articles which evolved into the US National Library of Medicine (NLM). This index was supplanted by PubMed (also a NLM project) and ceased publication in 2004. The largest database searched by PubMed is MEDLINE [3]. Citations in MEDLINE are collected from 5,582 (as of Nov 2011) [7] medical journals and it has 19,455,996 total records (as of Feb 2012) [6] from 1966 to the present with articles added to MEDLINE at the average rate of over 2000/day [6]. Each of these articles have been manually indexed by one of 100 human domain experts with MeSH terminology, 699,420 were indexed in 2010 [5]. Since PubMed searches MEDLINE and other resources, it is a little larger: it has 21,578,070 total records (as of Feb 2012) [6] from 1949 to the present. Other sources it searches are, for example: (1) the 460,488 [6] articles not yet indexed with MeSH terminology, but are waiting to be processed (i.e. indexed with MeSH) into the MEDLINE system, and (2) the 519,953 [6] records from OLDMEDLINE which contains records from the years 1946 to 1965. It is free to search MEDLINE and PubMed, and popular: it was directly searched 1.8 billion times in 2011 [5], an increase of 13% over the 1.6 billion searches in 2010.

The Medical Subject Headings (MeSH) are used as the basis for the navigation structure in PifMed. Updated yearly, the MeSH taxonomy was specifically designed by medical librarians for the organization of medical literature. With 16 main category headings at the root, and as deep as 11 levels, the 26,142 descriptors in MeSH 2011 [8] provide a thorough categorization, ideal for our task. Indexers can select one or a few specific MeSH descriptors associated with an article as a `Major Topic`. PifMed uses this feature of the data to give users the option to narrow (i.e. an article is added to a category node only if that category is marked as a `Major Topic` of the article) or widen (i.e. category nodes include every article associated with that category) the search.

## 5. EXPERIMENT

The study design has two components, paired use-time analysis to determine efficiency and a questionnaire to evaluate usability. The two systems tested PubMed and PifMed where both used to search MEDLINE only and filter out articles without abstracts. PubMed presents the articles in a ranked list (20 per page) and PifMed presents the articles in a browsable tree of hierarchically organized categories: exact same article set, different presentation.

### *Research Question.*

*In general, is the proposed navigation structure more effective than ranked lists for MEDLINE?*

### *Task.*

First, participants were shown a short PowerPoint presentation demonstrating each system. Then participants were asked to enter a query of their choosing into System A and find an article of interest. Then participants were asked to enter the exact same query into System B and find the same article or one equally or more interesting. Each user did 3 queries, then filled out a short, 3-part questionnaire. Participants were given a small monetary compensation for their time. All testing took place in a secure, quiet, well-lit, graduate computer science lab at Dalhousie University. The use-time data for each participant query is recorded (in seconds) as a pair (e.g. `'diabetes'[PifMed(91), PubMed(127)]`).

All factors have been *equalized*, *randomized* or *isolated* for this study. The within-group experimental design equalizes many possible confounding factors since the same participant uses each system immediately after the other. So the sense of relevancy is duplicated, the query terms are the same, the returned article set is identical, the time of day, user age, gender, and other general attributes were identical. To eliminate order effects, the order of systems is randomized for every query. The only difference is the interface, so any differences can be attributed to that isolated factor with a high degree of confidence. These pairs are then evaluated using a paired $t$-test.

With order effects randomized, the two common difficulties with within-groups study designs remain: *fatigue* and *learning effects* [4]. Fatigue is not an issue since no participant spent more than 40 minutes to complete study. The learning effect was largely uncontrolled. However, for users unfamiliar with both systems, it would have an equal (thus negligible) effect. For participants familiar with PubMed, this effect would be a handicap to the test system, thus should the reader believe the learning effect to be a significant factor for this study design, they should see chance of Type B error (false negative) as higher than usual.

### *Population.*

Each user rates (from 1 to 7) their own familiarity with 9 aspects: Computers, Computer Search, MeSH, PubMed, Medicine, Biology, Psychology, Health Informatics and Bioinformatics. We use these aspects to identify populations within the pool of users. We have two sets: *General Users* and *Target Users*. The profiles of these sets are shown in **Table 1**. A *General User* is any user scoring at least a 5 in both Computers and Computers Search which we feel are the key competencies. For this study all 27 participants tested fit this profile. A *Target User* is a user scoring at least a 6 in Computers and Computer Search and at least a 4 in one of the following PubMed, Medicine, Biology, Psychology, Health Informatics and Bioinformatics. For this study 22 participants of the 27 participants tested, fit this profile.

| General | Target | Familiarity with... |
|---|---|---|
| 7 | 7 | Computers |
| 7 | 7 | Computer Search |
| 2 | 2.5 | MeSH |
| 3 | 4.5 | PubMed |
| 4 | 5 | Max(PubMed, Medicine, Biology, Psychology, Health/Bioinformatics) |

**Table 1: The self-assessed knowledge ratings of our participants was used to separate out the *Target User* sub-group.**

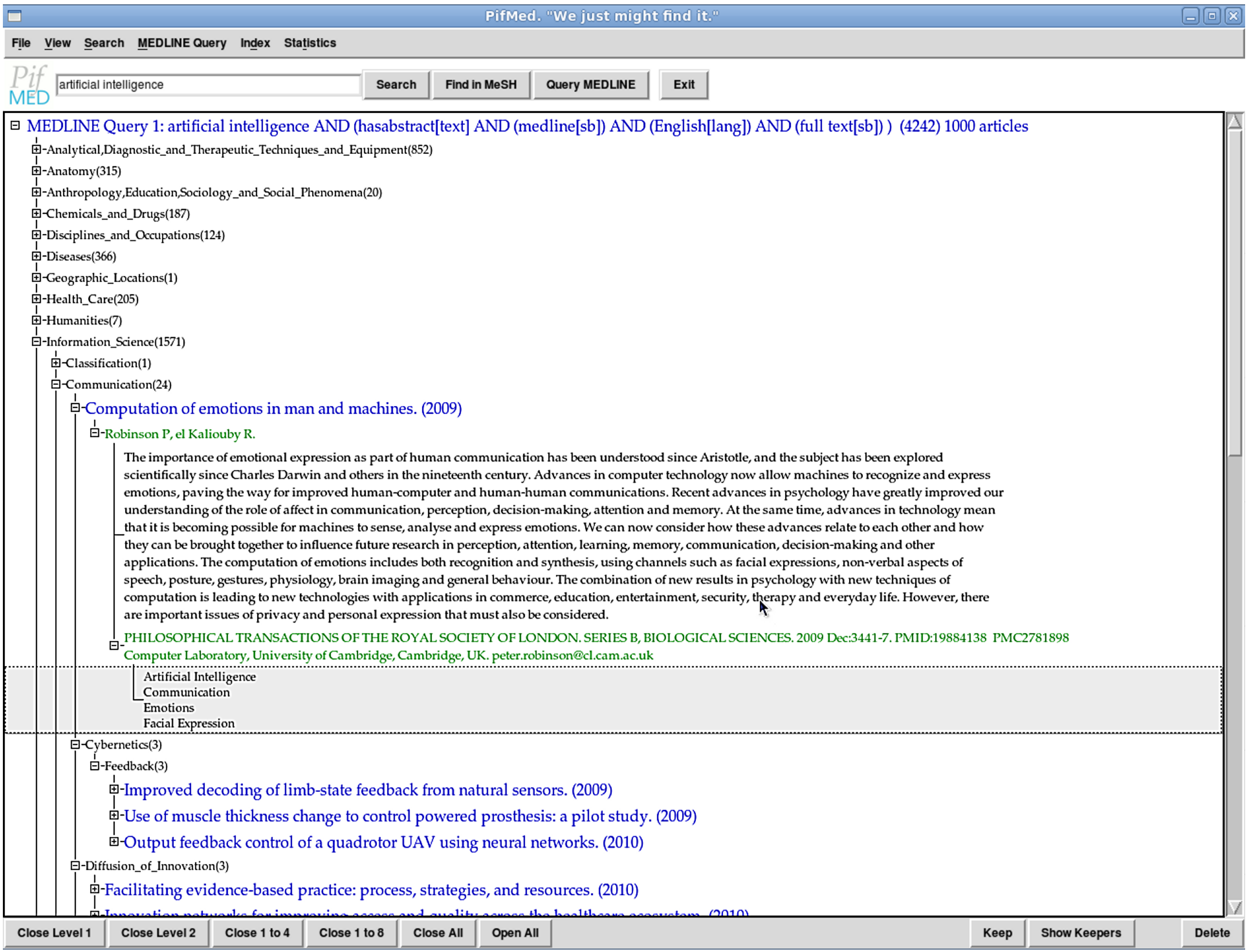


**Figure 1: The PifMed user interface. The root of the tree is open showing 10 MeSH main categories, each followed with a number which indicates how many articles are in each sub-tree, the 10th node is open. We see a fully open article node with the title '*Computation of Emotions in...*' under the main heading of Information Science and sub-heading of Communication. The authors, abstract and journal are revealed as are the four highlighted MeSH categories which are also associated with this article. This article is duplicated into each of these four category nodes on the tree.**

# 6. RESULTS

After the search tasks were complete the user was asked to fill out a questionnaire. Part II of that questionnaire is the basis for our determination of the usability of PifMed alone and the usability of PifMed in comparison to PubMed. **Table 2** shows the results of this part of the questionnaire for the *General User* population and for the *Target Users*. Each question in **Table 2** has two rows. The first row displays the results for PifMed alone and the second row displays the results for the comparison of PifMed and PubMed.

The reader will notice, each question in **Table 2** has an aspect in bold before each question. These aspects are known as the '5 Es' of usability: Effectiveness, Efficiency, Engagement, Error Tolerance, and Easy-to-Learn [10]. Each instance asks a question to evaluate one of those aspects. There are 2 questions for each aspect. The final question asks for an overall rating.

Each of these aspects can be measured as the median score of all ratings of questions of a that aspect. For example, the 4 ratings for "**Effectiveness**: *Did PifMed give you relevant results?*" and "**Effectiveness**: *Did PifMed help you make up your mind on what you were looking for?*" from each user can be averaged to give PifMed an overall rating for Effectiveness, **Table 3** shows the median results for each of the populations.

## 6.1 Statistical Conclusions

These tests show that participants find articles in large result sets faster with a PifMed than with PubMed. The mean use-time for PifMed is always faster in the final analysis, no matter how the data has been partitioned, but PifMed is strongest when short search result sets are removed, for both populations ($p = 0.0013$ & $p = 0.0113$). Overall, this set of participants were an average $\sim$26 seconds faster with a PifMed than with PubMed.

Furthermore, if speed was negligible, PifMed is still the more usable, since all users rated PifMed as more usable than PubMed. The analysis of the results show PifMed to be faster and a better user experience for this task.

# 7. LIMITATIONS

This system is meant for navigating large result sets, thus very short result sets can be seen as a limitation. There is no doubt ranked lists are a more effective means of displaying

| General | Target | Questions |
|---|---|---|
| 6<br>5 | 6<br>5 | **Effectiveness**: Did PifMed give you relevant results? |
| 4<br>4 | 4<br>4 | **Efficiency**: Did PifMed respond quickly? |
| 6<br>6 | 6<br>6 | **Engaging**: Did PifMed encourage you to explore the results? |
| 7<br>4 | 7<br>4 | **Error Tolerance**: Did you notice any errors in PifMed? |
| 7<br>5 | 7<br>5 | **Easy to Learn**: Was PifMed easy to learn and understand? |
| 6<br>6 | 5<br>5 | **Effectiveness**: Did PifMed help you make up your mind on what you were looking for? |
| 6<br>5 | 6<br>4 | **Error Tolerance**: How would you rank your confidence in the results? |
| 6<br>5 | 6<br>4 | **Easy to Learn**: Do you have a good understanding of the capabilities of PifMed? |
| 6<br>5 | 6<br>5 | **Efficiency**: Rate the ease (or difficulty) in retracing your steps. |
| 6<br>6 | 5.5<br>5 | **Engaging**: Did PifMed help you browse to interesting papers you did not expect? |
| 6<br>5 | 6<br>5 | **Overall**: Please rate the ease of using PifMed overall. |
| 6<br>5 | 6<br>5 | Median for **PifMed**<br>Median for **Comparison** |
| 6 | 5 | Median for **Overall Usability** |

**Table 2: The median results of Part II of the questionnaire. For each population there are two medians for each question. The first number is the median user rating for PifMed alone. The second number is the median user rating for the comparison of PifMed to PubMed.**

| General | Target | Usability Measure |
|---|---|---|
| 6 | 5 | Effective |
| 4.5 | 5 | Efficient |
| 6 | 6 | Engaging |
| 6 | 5.5 | Error Tolerant |
| 5.5 | 6 | Easy to Learn |
| **6** | **5** | **Total Average Usability** |

**Table 3: The final analysis of the qualitative results for the *General User* and *Target User* populations. Medians (on the left) of each of the measures of usability are shown and the overall usability is shown on the bottom line.**

20 or less results, these results can be seen to support this. This is an acceptable limitation since this is not the task PifMed is meant to tackle.

This system relies on the efforts of the NLM Indexers. One could see this dependency as a limitation: Should these indexers stop indexing, no new articles could be added to the tree. This is not a certainty. Many publishers (and authors) presently include suggested MeSH categorization meta-data with citation submissions to the NLM. Furthermore should the task of indexing be resumed by users, Wikipedia shows high-quality content can be publicly generated.

This system will only function within the confines of MEDLINE citations. This limitation to the medical domain is only a constraint of this implementation, not of the navigational structure itself. Should any other digital library, such as the ACM Portal (*http://portal.acm.org*), provide a deep and rigorous hierarchical categorization and make the system publicly available, this structure could be quickly adapted to another domain. In fact, preliminary tests were done on the ACM category hierarchy, but the hierarchy was too shallow to provide comparable results.

| Population | General Users | | Target Users | |
|---|---|---|---|---|
| Result Set Size | All | Long | All | Long |
| PubMed Mean | 100s | 106.62s | 95.79s | 102.23s |
| PifMed Mean | 82.17s | 80.18s | 83.56s | 78.32s |
| Mean Diff. | 17.83s | 26.44s | 12.23 | 23.91s |
| PubMed STD | 64.02s | 64.13s | 68.95s | 69.07s |
| PifMed STD | 42.78s | 41.59s | 43.36s | 41.53s |
| $p$-value | **0.0146** | **0.0013** | 0.0642 | **0.0113** |

**Table 4: A detailed summary of the statistical analysis. The Mean Differences and statistically significant $p$-values ($p$ <0.05) are in bold.**

# 8. FUTURE WORK

The mature, integrated knowledge-based systems developed at the NLM enable several avenues for future work. Through the Unified Medical Language System (UMLS) any MeSH descriptor can be used to access domain-specific NLP tools (SPECIALIST NLP System), a network of associative/ hierarchical semantic relationships (UMLS Semantic Network) and a mapping to +100 other categorization systems (UMLS Metathesaurus). For example, we could use the UMLS Metathesaurus to map the MeSH categorization into The Library of Congress classification for an alternate tree construction. Furthermore, we could take advantage of the UMLS Semantic Network to construct trees with associative relationship links and nodes (such as process of, part of, treats, disrupts) in tandem with the hierarchical tree to give users other meaningful pathways through which they may explore large result sets.

## 8.1 PifMed Web Version

Some of the limitations found in the user study have been addressed with the implementation of a second version of PifMed. **Figure 2** shows the a prototype web version of PifMed, which was built with a combination of HTML, XML, XSLT, Javascript, Perl and PHP, to address some of the limitations of the original Perl/Tk version. Specifically, lack of web version, status bar (built-in to web browser), interactive hyperlinked category names, improved node toggling and full text retrieval. Being web-based, the foundation was coded in HTML, the browsable tree was implemented in Javascript, the back-end search engine remains in Perl, the articles are returned in XML and rendered with XSLT and the whole system is glued together with PHP.

# 9. CONCLUSION

We have shown users prefer PifMed to PubMed in terms of usability, demonstrated by the analysis of Part II of the questionnaire. It has also been shown that users browsed to interesting results faster with PifMed than with PubMed. Paired $t$-tests have shown this statistically significant result to have improved significance when small results sets are partitioned out of the data set. This finding strengthens the evidence supporting my hypothesis that use of the hierarchical navigation structure is more effective and efficient then the conventional method for browsing large result sets.

Its success with large result sets is a key finding for 3 reasons:

1. **Query Expansion**. Query expansion returns increases the size of result sets. For query expansion to be viable we need a way for users to adequately navigate those results, PifMed is such a way.

2. **Growth of Corpus**. There is every indication MED-

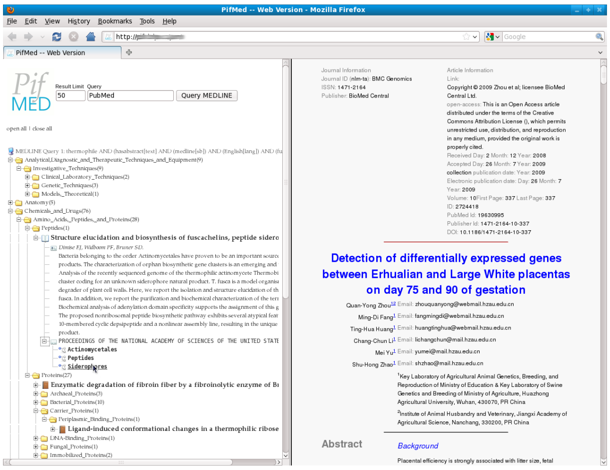


**Figure 2: The web version of PifMed has a browsable tree shown in the left frame and an article, rendered into HTML from XML, on the right.**

LINE will continue to grow, thus result sets for any given query will continue to grow. However, the rate of growth within Categories growth will be slower, that is, ∼50,000 new articles/week will be categorized across 26,000 categories. The hierarchical categorization mitigates the rate of growth by localizing growth into categories.

3. **Longevity of Queries** Changes in result sets will be focused into categories instead of drastically reorganizing ranked lists. Since no matter how many articles are added, old articles will always remain in their original category: PifMed behaves predictably over time.

The reliance on the MeSH taxonomy was a major concern. Since it is the foundation of this particular implementation of this browsing model, if users rejected it as navigational structure, the whole project would fail. However, these results show that users accepted and quickly adapted to this largely unfamiliar taxonomy. Evidence of this is shown by the very few complaints about MeSH in the written comments or verbal comments. But the strongest support for the choice to base the navigation system on MeSH comes in the plain fact that users used it quickly and effectively to navigate to interesting results using only MeSH as their guide, with *no previous training or experience with the taxonomy*.

## 10. ACKNOWLEDGMENTS

The authors would like to acknowledge support from the NSERC CRD grant on Finding best evidence for evidence-based best practice recommendations in health care. The authors would also like to thank anonymous reviewers for comments.

## 11. REFERENCES

[1] T. Berners-Lee, O. Lassila, and J. Hendler. The semantic web. *Scientific American*, pages 34–43, May 2001.

[2] C. Carpineto, S. Osiński, G. Romano, and D. Weiss. A survey of web clustering engines. *Computing Surveys (CSUR)*, 41(3):1–38, Jul 2009.

[3] L. Hunter and K. B. Cohen. Biomedical language processing: What's beyond pubmed? *Molecular Cell*, 21:589–594, Mar. 2006.

[4] J. Lazar, J. H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction*. Wiley, Jan. 2010.

[5] N. L. of Medicine. Key medline indicators, Dec. 2009. http://www.nlm.nih.gov/bsd/bsd_key.html. Last access 2012.

[6] N. L. of Medicine. Data, news and update information, Jan. 2010. http://www.nlm.nih.gov/bsd/revup/revup_pub.html#med_update. Last Visited 2012.

[7] N. L. of Medicine. Number of titles currently indexed., July 2010. http://www.nlm.nih.gov/bsd/num_titles.html. Last Visited 2012.

[8] N. L. of Medicine. Fact sheet – MeSH, Jan. 2011. http://www.nlm.nih.gov/pubs/factsheets/mesh.html. Last access 2012.

[9] O. D. Project. ODP - open directory project. http://www.dmoz.org/, 2010.

[10] D. L. Stone, C. Jarrett, M. Woodroffe, and S. Minocha. *User Interface Design and Evaluation*. Morgan Kaufmann, 1st edition, 2005.

[11] US Government. Patent classification home page - internet. http://www.uspto.gov/web/patents/classification/index.htm, Sept. 2010.

[12] W. Wiegand. The "amherst method": The origins of the dewey decimal classification scheme. *Libraries & Culture*, 33(2):175–194, Apr 1998.

[13] Wikipedia. Categories. http://en.wikipedia.org/wiki/Portal:Contents/Categories, 2010.

[14] Yahoo! Yahoo! answers. http://ca.answers.yahoo.com/dir/index, 2010.