Evaluation of a semi-automatic indexing tool for the biomedical literature using semantic similarity in MeSH

Sandy D. Tao
National Library of Medicine
Associate Fellow, 2004-2005

Aug. 15, 2005

Project Leader

Olivier Bodenreider

Table of Contents

1. Introduction

The purpose of this project is to compare human indexing (gold standard) to the output of a semi-automatic indexing tool in order to determine the degree of agreement between the systems by using semantic similarity and identity match methods. It is also to describe the characteristics of the performance of the semi-automatic indexing program, and to identify ways in which the semi-automatic indexing program can be improved.

2. Background

The Indexing Initiative project has been developed at NLM since 1996 and aims at suggesting MeSH main headings for MEDLINE citations. The output of the semi-automatic indexing process is a list of suggested MeSH main headings. The indexing terms suggested Indexing Initiative system are made available to the indexers through the DCMS system and the feedback provided by the indexers contributes to evaluating the Indexing Initiative system.

In 2002, the Indexing Initiative system was evaluated by the core team of the NLM's Indexing Initiative (Indexing Initiative Project, 2002). The team compared the MeSH main headings suggested by the system to those assigned by the indexers (gold standard) using identity match method. While this evaluation identified several areas in which the semi-automatic indexing program can be improved, we feel such identity match evaluation may be too exclusive, as no credit is given to MeSH headings close but not identical to those assigned (e.g., a child for a parent).

We proposed to explore an alternative evaluation method based on the semantic similarity between the two sets of main headings obtained for a given citation (suggested by the system and assigned by the indexers). Existing techniques developed for comparing the semantic similarity of terms in a taxonomy were applied to this task.

3. Definitions

3.1 Medical Subject Headings (MeSH)

Medical Subject Headings (MeSH) is the National Library of Medicine's controlled vocabulary thesaurus. The MeSH thesaurus is used by NLM for indexing articles from 4,800 of the world's leading biomedical journals for the MEDLINE database.

It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. There are 22,997 concepts in MeSH. In addition to these headings, there are more than 151,000 headings called Supplementary Concept Records (formerly Supplementary Chemical Records) within a separate thesaurus. There are also thousands of cross-references that assist in finding the most appropriate MeSH Heading (MeSH, 2005).

In this experiment, the 22,997 concepts in 2004 MeSH are the basis for building the semantic similarity matrix for our evaluation.

3.2 Medical Text Indexer (MTI)

The Indexing Initiative team from NLM created a semi-automatic indexing tool, Medical Text Indexer (MTI) in 2000. MTI generates a list of recommended terms for human indexers to assist in their indexing efforts. The human indexers have the option of using any or all of the MTI recommended terms while indexing an article.

The MTI system consists of software for discovering MeSH headings for citation titles. Some of the major components of MTI include a MetaMap-based indexing method, the PubMed Related Citation algorithm, and Restricted to MeSH (Aronson, 2000).

We should note the human indexers may have been influenced by MTI during their indexing process. Joe Thomas (BSD) observed the new indexers frequently reference MTI when they begin using the system. However, once they become familiar with the MeSH concepts, they are less likely to depend on MTI for suggested terms. Furthermore, when the indexers encounter an article they are not familiar with, they may check with MTI to see what the system would suggest. It serves as an entry point for them. More discussion with Joe Thomas is attached in Appendix A.

3.3 Semantic Similarity

Semantic similarity refers to similarity between two concepts in a taxonomy such as MeSH (Lin, 1998). For this evaluation, we define semantic similarity as an assigned metric between two concepts based on the likeness of their meaning and their semantic content.

One of the ways to evaluate semantic similarity in a taxonomy is to evaluate the distance between the nodes corresponding to the items being compared. The shorter the path from one note to another, the more similar they are. However, this approach relies on the notion links in the taxonomy represent uniform distance (Rada et al., 1989).

Unfortunately, uniform link distance is much difficult to define and to control. An alternative way to evaluate semantic similarity is based on the notion of information content. The more information two concepts share in common, the more similar they are (Resnik, 1996).

MeSH is a poly-hierarchical vocabulary with various degrees of linking distance between the terms. Information content measurement provides a neutral way of constructing a static knowledge structure to the vocabulary and its multiple contexts.

In this study, the semantic similarity measure is based on the information content of 2004 MeSH vocabulary obtained from a 28-year corpus of MEDLINE database (1965-2003). This measurement is consistent with the information theory, and it compensates the heterogeneity in MeSH (Bodenreider, 2004).

4. Methodology

4.1 Sample Population

In order to compare human indexing to the output of a semi-automatic indexing tool, the MeSH indexes (Major Headings) were gathered on one year of 2004 MEDLINE

citations from the outputs of MTI recommendations and Medline gold standard (human indexing).

For 2004, Medline human indexers processed 901,442 unique PMIDs (PubMed IDs); all terms were mapped to 2004 MeSH.

For semi-automatic indexing, only a subset of MEDLINE citations was processed by MTI. This subset contained 418,223 unique PMIDs, of which 6153 PMIDs had terms that were not mapped to 2004 MeSH. Since MTI employs PubMed Related Citation algorithm (which generates term that does not belong to MeSH), this explains why there are terms that couldn't be mapped to 2004 MeSH.

We have a common dataset of 412,070 unique PMIDs that we may use to compare MTI recommendations and Medline gold standard. For this experiment, 1% random sample are taken from the dataset; a total of 4120 PMIDs are used as our evaluation sample.

4.2 Complications

In the process of generating semantic similarity matrix of MeSH concepts, we come across several complications; more specifically, the weights of disproportional trees, cycling of MeSH concepts, and concepts appear in more than one tree. In the end, we decide to partition the 15 MeSH trees to minimize these complications.

4.2.1 Weights issues

MeSH concepts are organized in 15 trees. The trees are not an exhaustive classification of the subject matter. Their structure frequently represents a compromise among the views and needs of particular disciplines and users (MeSH, 2005). After we calculate the information content of MeSH concepts for each tree, it's clear some trees are much denser and have more information content than other trees. This creates 15 disproportional trees that need adjustment. This disproportion reflects a distortion of the information content of the biomedical literature.

MeSH Trees at a glance:

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Physical Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology and Food and Beverages [J]
11. Humanities [K]
12. Information Science [L]
13. Persons [M]
14. Health Care [N]
15. Geographic Locations [Z]

4.2.2 Appearance of a concept under multiple trees

MeSH has a poly-hierarchical structure which remits many challenges as we soon discovered. The concept may appear in as many places as may be appropriate. When a concept appears under multiple trees, do they have the same meaning? Or are they conglomerate of meanings? How should we assign a value to this information content?

We find there are over 6000 concepts that appear more than once under different trees in 2004 MeSH vocabulary.

4.2.3 Cycling

In addition, there is an unique phenomenon of cycling of concepts, i.e., concept A may be a child to concept B under one tree, at the same time, concept A is also the parent to concept B under another tree. We encountered one instance of such occurrence, which is detailed below. However, we have not investigated if there is other evidence of this phenomenon occurring in the MeSH vocabulary.

Humanities [K01]
Ethics [K01.316]
Morals [K01.316.630]

Humanities [K01]
Philosophy [K01.752]
Ethics [K01.752.256]
Morals [K01.752.256.547]

Behavior and Behavior Mechanisms [F01]
Psychology, Social [F01.829]
Morals [F01.829.500]
Ethics [F01.829.500.519]

As the above example has shown, Ethics is the parent to Morals under the K tree, but it becomes a child of Morals under the F tree. Also note both concepts appear in multiple trees at the same time.

4.2.4 Partition

The decision is made to partition these 15 MeSH trees so we have 15 virtual root nodes. This solves the problem of disproportional trees. Since each tree is now an individual entity, they are not compared to each other. When comparing semantic similarity value between concepts from different trees, they are assigned with a value of zero since they do not share a common root. Logically, this also makes sense. Since these concepts are from two different trees, they are not similar.

<u>Concepts from different trees:</u>

A Tree	B Tree
A1	B1

Semantic Similarity value between A1 and B1 = 0

Partition also solves the problem of cycling and appearance of single concept under multiple trees. When a concept is under two or more trees, we keep the value that gives us the highest information content. With the cycling, this rule also applies. By separating the trees and treating them individually, we are able to work around many of the specific characteristics of MeSH.

4.3 Semantic Similarity Matrix

4.3.1 Information Corpus

To prepare a semantic similarity matrix, we selected the 2004 MeSH vocabulary. We obtained the raw frequencies of the 2004 MeSH concepts in the literature from the 1965-2003 of MEDLINE, a corpus of near 30-years of biomedical literature abstracts.

4.3.2 The Matrix

As we illustrated earlier, there are 22,997 concepts in 2004 MeSH. The semantic similarity matrix of MeSH concepts encompasses 264,431,005 similarity values between all possible concept pairs.

Total number of semantic similarity score between all possible MeSH pairs = $C^2/2$

where C is the number of concept in MeSH

4.3.3 Lin's Method

Lin's method is chosen to measure semantic similarity in MeSH. One of the major characteristics of Lin's Method is that the values are within the range of zero to one. We chose Lin's method because the semantic similarity values are within the range of the identity match (zero or one); therefore, we have a base to compare the findings later on.

Formula of Lin method:

$$sim(c1, c2) = \frac{2 \times [\ln Pms(c1, c2)]}{\ln P(c1) + \ln P(c2)}$$

Where P(c1) is the probability of c1, (number of times c1 and any of its child, occurs in the corpus / total number of terms in the corpus)

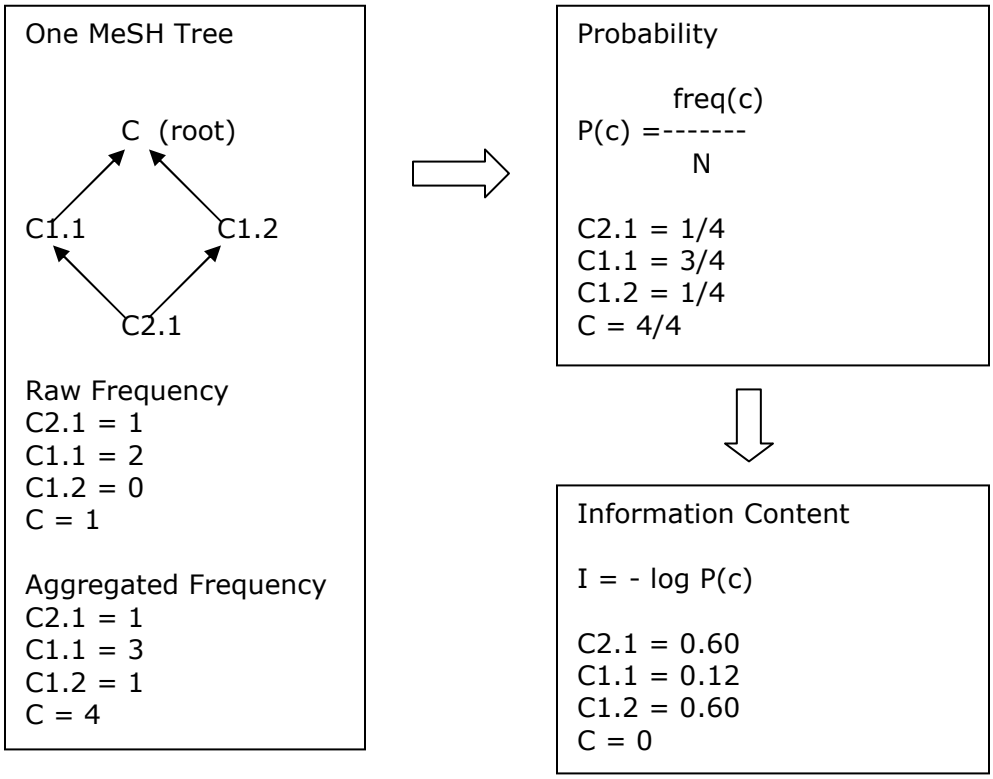
$$P(c) = \frac{freq(c)}{N}$$

Where Pms(c1, c2) is the probability of the minimum subsumer of c1, c2, i.e.

$$Pms(c1, c2) = \min_{c \in S(c1,c2)} \{P(c)\}$$

where S(c1, c2) is the set of parental concepts shared by both c1, c2.

4.3.4 An Example



As the example illustrates, the information content of a MeSH concept is quantified as inverse of the logarithm of the aggregated frequency. Thus, as the aggregated frequency of a concept increases, the information content of the concept decreases. The root node will have zero information content, where as the leaflets contain the most information content.

Since the more information two concepts share in common, the more similar they are (Resnik, 1996). The similarity value is based on the likeness, and how much information is shared between two concepts. A value of zero indicates two concepts are least alike and share no common information, a value of one indicates the concepts are most similar and share the most information content.

For instance, *Antigens, CD3* and *Antigens, CD* have a semantic similarity value of 0.76. *Antigens, CD3* and *Biological Markers* have a semantic similarity value of 0.68. From the MeSH tree, *Antigens, CD3* is closer to *Antigens, CD* than to *Biological Markers*.

Biological Markers [D24.185.101]
Antigens, Differentiation [D24.185.101.100]
Antigens, CD [D24.185.101.100.110]
Antigens, CD3 [D24.185.101.100.110.095]

4.4 Evaluation Process

4.4.1 Semantic Similarity

To generate the semantic similarity score for our sample, we take the bidirectional best match approach, i.e., best matches from MTI to Medline gold standard, and best matches from Medline gold standard to MTI.

For every PMID in our sample, each major heading generated by MTI is matched against major headings of Medline gold standard. The best match is chosen. This process is repeated on the opposite direction. All best matches will be accounted for, and aggregated to produce a single maximum average (see example in 4.6.3).

The semantic similarity value ranges from 0 to 1. Zero means the indexes generated by MTI are the least alike the indexes generated by Medline gold standard. One means they are the most similar.

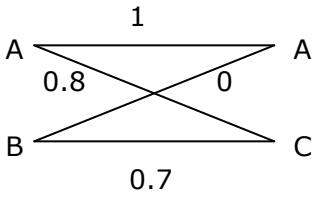
4.4.2 Identity Match

Similar approach is employed for identity match evaluation. The values for identity match are either 0 or 1. The terms are either identical, or they do not match at all.

4.4.3 Example

A simplified example for calculating the semantic similarity and identity match values are illustrated below. In this example, MTI generated 2 concepts A and B; Medline Gold Standard generated A and C. Note that A and B have a semantic similarity of zero. They are in separate trees under this scenario.

MTI Medline Gold Standard



Direction: From MTI to Medline Gold Standard

A-→ A 1
A-→ C 0

B-→ A 0.8
B-→ C 0.7

A-→ A 1

B-→ A 0.8

Direction: From Medline Gold Standard to MTI

A-→ A 1
A-→ B 0.8

C-→ A 0
C-→ B 0.7

A-→A 1

C-→B 0.7

$$\text{Semantic Similarity Score} \quad \frac{1 + 0.8 + 1 + 0.7}{4} = 0.875$$

Direction: From MTI to Medline Gold Standard

A-→1
B-→0

Direction: From MTI to Medline Gold Standard

A-→1
C-→0

$$\text{Identity Match Score} \quad \frac{1 + 0 + 1 + 0}{4} = 0.5$$

5. Results

5.1 Summary

Of the 4120 PMIDs sample, MTI and Medline gold standard indexing terms score an average of 0.53 based on the semantic similarity. As expected, with identity match, they have a 0.32 value, a lower average score.

On average, MTI generates 23.38 terms per PMID, while Medline gold standard generates 12.46 terms per PMID. They share a 6.03 common terms.

4120 PMIDS	Semantic Similarity	Identity Match	Shared IM	Medline Terms	MTI Terms	IM/Medline	IM/MTI
Average	0.53	0.32	6.03	12.46	23.38	0.50	0.26

5.2 Rule of Three

Medline gold standard follows the Rule of Three. The rule of three specifies that human indexers always try to select the most specific term that describes the concepts in an article. However, if more than three specific concepts treed under a more general concept common to all are discussed in an article, index the general concept (IM).

Here is an example when MTI did not follow this rule. MTI generated 3 narrower concepts instead of one single parent concept (Protein-Serine-Threonine Kinases).

PMID: 15187108

J Immunol. 2004 Jun 15;172(12):7324-34.
A novel role for p21-activated protein kinase 2 in T cell activation.
Chu PC, Wu J, Liao XC, Pardo J, Zhao H, Li C, Mendenhall MK, Pali E, Shen M, Yu S, Taylor VC, Aversa G, Molineaux S, Payan DG, Masuda ES.

From MTI to Medline:

Mitogen-Activated Protein Kinases(D020928) -> Protein-Serine-Threonine Kinases(D017346) 0.8607602
Mitogen-Activated Protein Kinase Kinases(D020929) -> Protein-Serine-Threonine Kinases(D017346) 0.7586931
MAP Kinase Kinase Kinases(D020930) -> Protein-Serine-Threonine Kinases(D017346) 0.75525683

Three children concepts from MTI all mapped into one single parent.

Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
- MAP Kinase Kinase Kinases [D08.811.913.696.620.682.700.559]
- Mitogen-Activated Protein Kinase Kinases [D08.811.913.696.620.682.700.565]
- Mitogen-Activated Protein Kinases [D08.811.913.696.620.682.700.567]

5.3 Concepts missed by MTI

Comparing the concepts in Medline Gold Standard to MTI indexing, we find that the check tags, location and time period information are routinely missed by MTI (see Appendix B).

We should note the performance of MTI is limited to what's on the text. If the check tags are not visibly present in the abstract, they may be difficult to pick up.

5.4 MTI's Tendency to index Narrower Concepts

After surveying a number of samples, we notice MTI's tendency to capture narrower concepts while Medline Gold Standard captures the broader concepts.

From the same example we demonstrated earlier: PMID 15187108, 15 best matches are found from the direction of Medline gold standard to MTI. Four matches have a perfect score of 1; one shows no match; the rest have various degrees of semantic similarity. The following five matches listed below demonstrate how MTI captures the narrower concepts while the Medline gold standard captures the broader concepts.

From Medline gold standard to MTI:

Antigens, CD(D015703) -> Antigens, CD3(D017252) 0.7605871
Antigens, CD [D24.185.101.100.110]
    Antigens, CD3 [D24.185.101.100.110.095]

Antigens, Differentiation, T-Lymphocyte(D000945) -> Antigens, CD3(D017252) 0.8748187
Antigens, Differentiation, T-Lymphocyte [D24.185.101.100.894]
    Antigens, CD3 [D24.185.101.100.894.095]

Biological Markers(D015415) -> Antigens, CD3(D017252) 0.6855663
Biological Markers [D24.185.101]
    Antigens, CD3 [D24.185.101.100.110.095]

Cell Line, Tumor(D045744) -> Jurkat Cells(D019169) 0.80381376
Cell Line, Tumor [A11.251.860.180]
    Jurkat Cells [A11.251.860.180.495]

DNA-Binding Proteins(D004268) -> NF-kappa B(D016328) 0.78245896
DNA-Binding Proteins [D12.776.260]
    NF-kappa B [D12.776.260.600]

5.5 Additional observations

Of the 4120 PMIDs, we sort the PMIDs based on their semantic similarity value. The top 50 (see Appendix C), has the highest values, indicates the best match scenario; the bottom 50 (see Appendix D), has the lowest values, indicates the worst match scenario. We examine the title and the publication type in hope it will give us an insight on why some articles are better matched while others are not.

In a closer examination, we speculate that MTI indexes well on reviews, research articles, case reports with a strong scientific emphasis. On the other hand, when it's foreign title, news articles, comments, editorial, and biography, MTI generally does a poor job than human indexers.

We should note, on the bottom 50 list, only 4 articles include abstracts. Since MTI indexes based on the title and abstracts, the lack of abstracts in specific publication types may also be a factor in the performance of MTI.

5.6 An Extended Example

Appendix E lists an extended example of PMID 15187108. In this example, MTI generated 22 Major Headings while Medline Gold Standard generated 15. Only 5 were exact identical matches. When applying the traditional metrics measure using identity match, the article scored 0.27. On the other hand, when semantic similarity method was applied, the article obtained a much higher score, 0.78.

A closer look at the concepts generated by MTI, the terms are the narrower counterpart of the terms from Medline Gold Standard. These two sets of indexes from MTI and Medline Gold Standard have a much higher degree of agreement than simple intersection would suggest.

Semantic similarity seems to provide a less-discriminatory evaluation of relatedness between lists of major headings than identity match method.

6. Discussion

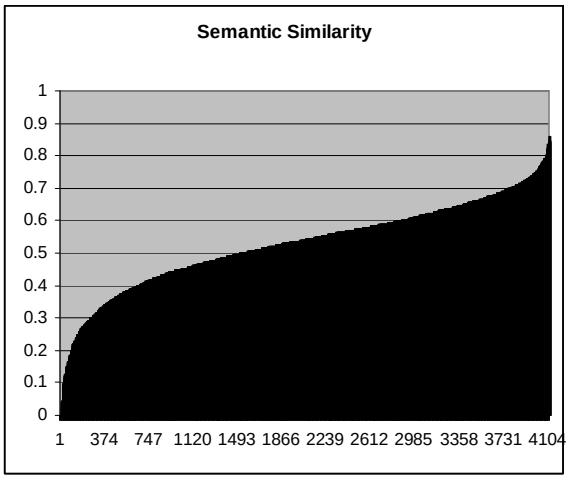
6.1 MTI

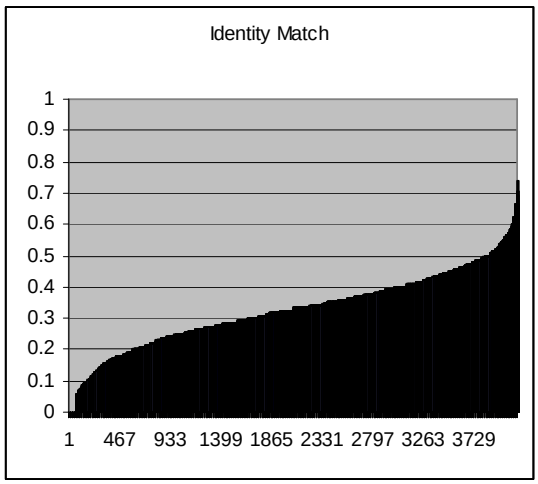
MTI methods of assigning indexing terms apply only to the title and abstract, while human indexers base their analysis on the full text of the article. This restriction caused some uneven performance of MTI, i.e., missing check tags, poor indexing for articles without abstracts. We suspect the performance of MTI can be improved when processing the full text articles instead of only abstracts.

One of the particular problems in processing natural language is with word sense ambiguity. Some of the worst indexed matches found in MTI contain publication type such as news articles, comments, and editorial. On the other hand, MTI's best indexed matches are on reviews, research articles, case reports with a strong scientific emphasis. We presume the texts on these specific scientific publications are less ambiguous, thus more comprehensible to MTI.

6.2 Semantic Similarity vs. Identity Match

The range for Identity Match scores is between 0 - 0.74, with an average of 0.32. The range for Semantic Similarity scores is between 0 – 0.86, with an average of 0.53. As illustrated in the graph below, semantic similarity measurement raised the average score higher; nonetheless, the distributions of the scores are similar in both methods.





Granted, semantic similarity measurement raises the score. What does it mean to have a higher score? Does it reveal a higher level of agreement among the terms generated by MTI and Medline Gold Standard?

It depends on the rationale behind the evaluation. For information retrieval purposes, it is important to evaluate how closely the two sets of terms generated using a semi-automated indexing vs. human indexing. If they are indeed semantically similar, it's plausible the effectiveness of information retrieval for both indexes is comparable.

On the other hand, if the purpose is to improve the semi-automated indexing, the evaluation using semantic similarity vs. identity match may not vary as much. Both methods provide substantial evidences and suggestions to improve MTI.

6.2.1 Identity Match

Does the identity match method give too little credit to the index terms that are close but not identical to those assigned? We believe the answer is yes. Identity match method is too restrictive. The close terms are excluded from the evaluation because they are not identical.

In PubMed, every MeSH term is automatically exploded. It retrieves citations that carry the specified MeSH heading and also retrieve citations that carry any of the more specific MeSH headings indented beneath it in the tree structure. For information retrieval purposes, identity match method does not give an accurate account of the relatedness/closeness of the terms generated between MTI and Medline human indexers.

To assist Indexers with a list of recommended terms from MTI, we think identity match evaluation method is effective and sufficient for that particular purpose. Since the human indexers make inference to the MTI suggested terms, the terms need to be precise and accurate. Recall should be limited; precision is what MTI should be aimed for.

6.2.2 Semantic Similarity

Does semantic similarity measure give too much credit to close index terms? Are we overly optimistic with a higher score using semantic similarity measures?

Yes, it's quite possible. Look at the following example. One of semantic similarity value assigned to the article is between *Brain Mapping* and *Image Enhancement*. It scores 0.10. What's the likelihood when someone is searching for brain mapping, it will be mapped to image enhancement?

They are semantically similar because both terms are under the E tree, and belong to *Analytical, Diagnostic and Therapeutic Techniques and Equipment*. *Image Enhancement* describes the imaging used in diagnostic procedure; *Brain Mapping* describes a particular investigative technique.

It's unlikely these two concepts have a factual relatedness. This emphasizes a need for re-examination of the semantic similarity measurement. For future study, we propose to rank the semantic similarity scores and eliminate any score that's lower than a specific number (such as 0.10). That bear minimum would be the lowest threshold for any terms to be a significant addition to the overall score.

PMID: 15376887
IEEE Trans Pattern Anal Mach Intell. 2004 Mar;26(3):408-13.
Strong Markov random field model.
Paget R.

Brain Mapping (D001931) -> Image Enhancement(D007089) 0.103726596

Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
Diagnosis [E01]
Diagnostic Techniques and Procedures [E01.370]
Diagnostic Imaging [E01.370.350]
Photography [E01.370.350.600]
Image Enhancement [E01.370.350.600.350]
Investigative Techniques [E05]
Brain Mapping [E05.132]

7. Suggestions for Future Works

Since this is a preliminary study using semantic similarity measure, we hope to refine the evaluation process in the future. We propose to test on a larger sample size, to re-compute the sample without the check tags, to set a minimum threshold for evaluation.

We also suggest implementing the Rule of Three on the output of MTI. When more than three specific concepts treed under a more general concept in an article, MTI will index the general concept (IM).

Lastly, we suggest conducting a retrieval experiment to compare the indexes from MTI and Medline human indexers. If the two sets of terms generated using a semi-automated indexing vs. human indexing are in fact similar, it's plausible the effectiveness of information retrieval for both indexes is comparable. The retrieval experiment will be able to test the hypothesis.

Acknowledgements

I would like to acknowledge the following people who have assisted me in this project. I would not be able to complete the project without the help of these wonderful folks!

Olivier Bodenreider - LHC
Kelly Zeng - LHC
Joe Thomas - BSD
Cliff Gay - LHC
Barbara Rapp - LO

Appendix A: Human Indexing

MTI generates a list of recommended terms for human indexers to assist in their indexing efforts. I spent some time discussing the usage of MTI by the indexers with Joe Thomas (BSD).

Here is a list of items he observed in MTI:

- MTI generates misleading terms
- Terms are either too general (lack specificity) or too specific. The depth is uneven.
- Unable to distinguish important terms (IM) from non-important ones
- MTI generates too many terms. In average, human indexers would prefer 10 good terms that cover the depth of the article.
- MTI does not have any ranking of the terms it suggested.

Other observations:

- MTI does a fairly good job when indexing pre-clinical, scientific publication with descriptive titles.
- The new indexers frequently reference MTI when they first start using the system. However, once they become familiar with the MeSH concepts, they are less likely to depend on MTI for suggested terms.
- When the indexer encounters an article they are not familiar with, they would check with MTI to see what the system would suggest. It serves as an entry point for them.

Suggestions he hopes to include in MTI:

- Suggestions for Gene Link
- Suggestions for Check Tags, chemical terms, and entry terms
- Ranking of the suggested terms

Appendix B: Example of Concepts Missed by MTI

Term in Medline (not matched in MTI)	Number it missed
Support, Non-U.S. Gov't	1476
Human	1117
Male	1046
Female	858
Adult	652
Middle Aged	614
Comparative Study	566
Support, U.S. Gov't, P.H.S.	489
Aged	412
English Abstract	271
Animals	226
Time Factors	196
United States	163
Adolescent	158
Support, U.S. Gov't, Non-P.H.S.	141
Aged, 80 and over	121
Child	104
Dose-Response Relationship, Drug	87
Child, Preschool	54
Treatment Outcome	48
Molecular Sequence Data	45
Disease Models, Animal	45
Models, Biological	43
Mice	42
Mutation	38
Great Britain	38
In Vitro	35
History of Medicine, 20th Cent.	32
Reference Values	32
Infant	31
Diagnosis, Differential	30
Infant, Newborn	27
Models, Molecular	26
Prognosis	25
Blotting, Western	23
Pregnancy	23
Reverse Transcriptase Polymerase Chain Reaction	22
Escherichia coli	21
RNA, Messenger	21
Rats	19
Immunohistochemistry	19
Risk Factors	18
Drug Therapy, Combination	18
Chronic Disease	17
Species Specificity	17
England	16
Follow-Up Studies	16
Recurrence	16
Cell Line	16
Analysis of Variance	16
Age Factors	16
Drug Interactions	15
Acute Disease	15

Phenotype	15
Computer Simulation	15
Plasmids	14
Forecasting	14
Risk Assessment	14
Canada	14
History of Medicine, 19th Cent.	13
Disease Progression	13
Sensitivity and Specificity	12
Precipitin Tests	12
Brain	12
Kinetics	12
Aging	12
Germany	12
DNA	12
Severity of Illness Index	12
Cells, Cultured	11
France	11
Australia	10
Hydrogen-Ion Concentration	10
Temperature	10
Italy	10
Pyrimidines	10
Quality of Life	10
Liver	10
Patient Selection	10
Cell Differentiation	10
Retrospective Studies	10
Kidney	10
Algorithms	10
Gene Expression Regulation	10
Clinical Trials	10
Models, Theoretical	10
Fatal Outcome	10
Amino Acid Sequence	10
Europe	10
Terminology	10
Drug Synergism	10
In Situ Hybridization	10
Research Design	9
Gene Expression	9
Case-Control Studies	9
Injections, Intraventricular	9
Anti-Bacterial Agents	9
Image Processing, Computer-Assisted	9
Practice Guidelines	9

Appendix C: The Top 50 Best Matches

PMID	Title	Publication Type	
14748625	Complicated acute aortic dissection type B caused by femoral cannulation for endoscopic coronary artery bypass surgery.	Case Reports	
14677609	Restoration of endodontically treated teeth without posts.	Case Reports	
15070415	Ureterolithiasis after Cohen re-implantation--case report	Case Reports	
15289145	Decrease of blue cone sensitivity in acute idiopathic blind spot enlargement syndrome	Case Reports	
15262441	Unbalanced t(2;19) and t(2;16) in a neurofibroma	Case Reports	Letter
15354671	Oral and maxillofacial pathology case of the month. Mucoepidermoid carcinoma	Case Reports	
15466374	MMP-12, an old enzyme plays a new role in the pathogenesis of rheumatoid arthritis?	Comment	
15112673	Misuse of terminology to imply that 1,25-dihydroxy-vitamin D is a nutrient: there is no evidence for an association between vitamin D and allergy	Comment	letter
14665543	Is there any alternative to the Bispectral Index Monitor?	Editorial	
15121206	A statistical method for evaluation quality of medical images: a case study in bit discarding and image compression.	Evaluation Studies	
15049367	Fractionation of Trichoderma reesei cellulases by hydrophobic interaction chromatography on phenyl-sepharose.	Evaluation Studies	
15376887	Strong Markov random field model	Evaluation Studies	Validation Studies
14970791	Ocular torsion: rotations around the "WHY" axis.	Lectures	
15232514	New perspective on the management of hyperlipidemi	Letter	
14719492	Senate committee calls for major new spending on health care	Newspaper Article	
15328417	Multiple pathways process stalled replication forks	Review	Review, Tutorial
14630040	The discovery, synthesis, and role of pyridoxal phosphate: phase I of many phases in the Gunsalus odyssey.	Review	Review, Tutorial
14504459	Getting in the ring: proline-directed substrate specificity in the cell cycle proteins Cdc14 and CDK2-cyclinA3.	Review	Review, Tutorial
14708430	Developing self-evaluation skills: a pragmatic research-based approach for complex areas of nursing.	Review	Review, Tutorial
14728005	Clopidogrel: potential in the prevention of cardiovascular events in patients with acute coronary syndromes.	Review	Review, Tutorial
15085367	Second-generation real-time three-dimensional echocardiography. Finally on its way into clinical cardiology?	Review	Review, Tutorial
15055442	Emerging tumor entities and variants of CNS neoplasms	Review, Tutorial	Review
14967138	The RITS complex-A direct link between small RNA and heterochromatin	Review, Tutorial	Review
15451242	Stress distributions in adhesively cemented ceramic and resin-composite	Validation Studies	

	Class II inlay restorations: a 3D-FEA study.	
14655982	The moving dynamic random dot stereosize test: validity and repeatability	Validation Studies
15267047	Prevalence of HIV-positives and hepatitis B surface antigen-positives among donors in the University of Benin Teaching Hospital, Nigeria.	Research Article
15090179	Development of an assay suitable for high-throughput screening to measure matrix metalloprotease activity.	Research Article
14674269	Use of limited proteolysis to identify protein domains suitable for structural analysis.	Research Article
15086637	Violence in the care of adult persons with intellectual disabilities.	Research Article
15144978	High throughput screening of library compounds against an oligonucleotide substructure of an RNA target.	Research Article
15107066	Identification of a Fusobacterium nucleatum 65 kDa serine protease.	Research Article
14679902	[Membrane-bound forms of serine proteases of Bacillus intermedius]	Research Article
14729982	c-Kit-mediated overlapping and unique functional and biochemical outcomes via diverse signaling pathways.	Research Article
15065076	Matrix metallo-proteinase (MMP-2) organoboronate inhibitors.	Research Article
15156112	Detraining effects on the mechanical properties and morphology of rat tibiae.	Research Article
15162556	Got mold? Hospitals make progress in the fight against fungus.	Research Article
15319012	Are "carve outs" in or out?	Research Article
14712229	G-protein-coupled receptor-mediated activation of rap GTPases: characterization of a novel Galphai regulated pathway.	Research Article
14679588	Metaphyseal chondrodysplasia with cone-shaped epiphyses: a specific form involving the lower limbs	Research Article
15218668	Changes in thought for dental hygienists	Research Article
15331324	Role of the polypeptide region of a 33kDa mycobacterial lipoprotein for efficient IL-12 production	Research Article
14585725	Water sorption characteristics of light-cured dental resins and composites based on Bis-EMA/PCDMA	Research Article
14978741	Activity of the matrix metalloproteinase-9 promoter in human normal and tumor cells	Research Article
15225620	Activation of Ca2+/calmodulin-dependent protein kinase II is involved in hyperosmotic induction of the human taurine transporter.	Research Article
15326477	P38SAPK2 phosphorylates cyclin D3 at Thr-283 and targets it for proteasomal degradation.	Research Article
15344582	Simplified ceramic restorations using CAD/CAM technologies.	Research Article
14710641	[Structure and limitations of German hospitals with regard to the future--the example of the Vivantes Group]	Research Article
14652992	Turning Medicare billings into revenue	Research Article
14981846	Rising medical, admin costs push 2004 premiums higher.	Research Article
14978098	A regulatory role for CD37 in T cell proliferation	Research Article

Appendix D: The 50 Worst Matches

PMID	Title	publication type		Abstracts
14667838	The RGD story: a personal account.	Biography	Historical Article	
14677495	Sunter throws his cap into SAMA's ring.	Biography	Historical Article	
15045993	Criticism of authority in the writings of Moses Maimonides and Fakhr Al-Din Al-Razi.	Biography	Historical Article	yes
14631414	Stoned 1.	Case Reports	Case Reports	
15159712	Images in emergency medicine.	Case Reports		
14661308	Generalized vaccinia.	Case Reports		
14661613	[Arteriovenous malformations mimicking dilated medulla oblongata veins in Sturge Weber syndrome]	Case Reports		
15116707	[Infected urachal cyst]	Case Reports	Letter	
14646828	Early defervescence and SARS recovery.	Case Reports		yes
15353160	[Optic disk drusen: what are the advantages of the new imaging techniques?]	Case Reports		
15156476	Excoriations and ulcers on the arms and legs.	Comment	Editorial	
15060544	Getting beyond diagnostic accuracy: moving toward approaches that can be used in practice.	Comment	Letter	
15306784	The vector that got away.	Comment	Letter	
15202434	Golden rule of economics yet to strike prospectors.	Comment	Editorial	
14735639	Think STOP before going "off-label".	Comment	Letter	
15315289	Professionalism.	Comment		
14669777	Sticking the landing: how to create a clean end to a medical visit.	Comment	Letter	
15014316	The lack of science behind the standard of care.	Comment	Letter	
12743222	Correlation between bacteriologic eradication and clinical cure in acute otitis media.	Comment	Editorial	
15241388	Bringing epilepsy out of the shadows.	Comment	Letter	
15054311	Fat chance of measuring food intake accurately.	Congresses		
14740643	[Impression and discussion on RSNA'03 (discussion)]	Congresses		
15287694	German Chemical Society (GDCh) biannual conference in Munich 2003.	Congresses		
15057636	["Targets, drugs and carriers--novel therapeutic approaches"] Frontiers in Medicinal Chemistry--Annual Meeting 15-17 March 2004 Erlangen, Germany.	Congresses		
15039882	Auricular vision!	Editorial		
14770352	[Threshold]	Editorial		
15029056	[Digestive and visceral surgeons: an endangered species?]	Editorial		
14960385	How to show that an ineffective therapy works.	Editorial		
14753172	[Good fortune and eyeglasses]	Editorial		
15006081	View from the frontline.	Editorial		
15095110	[Editorial. Medical specialty regulations]	Editorial		
15108709	15th january 1913-The day pharmacy in Britain entered a new era.	Historical Article		
15304446	piece of my mind. For the obscure researcher.	Historical Article		
15198015	Equal treatment. Interview by Terry Philpot.	Interview		
15000020	The NMC defends its stance on fees. Interview by Mahua Chatterjee.	Interview		
15146884	Eric W. Taylor, MB. Interview by Vicki	Interview		

	Glaser.			
15101459	PIC seeks new member for Board.	Letter		
15102656	On the definition of relevant disease.	Letter		
15332075	Reading ratios.	News		
14959542	Smallpox mixes make a stir.	News		
15274242	[HIV/HCV double infection: combination is a clear therapy option]	News		
14666598	Hey, we're all victims here.	News		
14755259	Model droplets.	News		
15164558	No time for wrinkles.	News		
15357469	Secrets behind the mask.	News		
15214114	Eight days a week?	News		
15168517	[Immunoregulatory role of the protein quality control system]	Review	Review, Tutorial	
14970896	[Progress in biotransformation of triptolides and bufadienolides]	Review	Review, Tutorial	yes
15165389	Photobiological basis and clinical role of low-intensity lasers in biology and medicine.	Review	Review, Tutorial	yes

Appendix E: An Extended Example

PMID: 15187108

J Immunol. 2004 Jun 15;172(12):7324-34.

A novel role for p21-activated protein kinase 2 in T cell activation.

Chu PC, Wu J, Liao XC, Pardo J, Zhao H, Li C, Mendenhall MK, Pali E, Shen M, Yu S, Taylor VC, Aversa G, Molineaux S, Payan DG, Masuda ES.

Rigel Inc., 1180 Veterans Boulevard, South San Francisco, CA 94080, USA.

To identify novel components of the TCR signaling pathway, a large-scale retroviral-based functional screen was performed using CD69 expression as a marker for T cell activation. In addition to known regulators, two truncated forms of p21-activated kinase 2 (PAK2), PAK2DeltaL(1-224) and PAK2DeltaS(1-113), both lacking the kinase domain, were isolated in the T cell screen. The PAK2 truncation, PAK2DeltaL, blocked Ag receptor-induced NFAT activation and TCR-mediated calcium flux in Jurkat T cells. However, it had minimal effect on PMA/ionomycin-induced CD69 up-regulation in Jurkat cells, on anti-IgM-mediated CD69 up-regulation in B cells, or on the migratory responses of resting T cells to chemoattractants. We show that PAK2 kinase activity is increased in response to TCR stimulation. Furthermore, a full-length kinase-inactive form of PAK2 blocked both TCR-induced CD69 up-regulation and NFAT activity in Jurkat cells, demonstrating that kinase activity is required for PAK2 function downstream of the TCR. We also generated a GFP-fused PAK2 truncation lacking the Cdc42/Rac interactive binding region domain, GFP-PAK2(83-149). We show that this construct binds directly to the kinase domain of PAK2 and inhibits anti-TCR-stimulated T cell activation. Finally, we demonstrate that, in primary T cells, dominant-negative PAK2 prevented anti-CD3/CD28-induced IL-2 production, and TCR-induced CD40 ligand expression, both key functions of activated T cells. Taken together, these results suggest a novel role for PAK2 as a positive regulator of T cell activation.

PMID: 15187108 [PubMed - indexed for MEDLINE]
MTI result: Protein Kinases|Receptors, Antigen, T-Cell|Antigens, CD28|Jurkat Cells|Protein-Tyrosine Kinase|T-Lymphocytes|Antigens, CD3|Ca(2+)-Calmodulin Dependent Protein Kinase|Signal Transduction|Mitogen-Activated Protein Kinases|Mitogen-Activated Protein Kinase Kinases|Enzyme Activation|Lymphocyte Specific Protein Tyrosine Kinase p56(lck)|Interleukin-2|Receptor-CD3 Complex, Antigen, T-Cell|MAP Kinase Kinase Kinases|Phosphorylation|Antigens, CD2|Cell Physiology|NF-kappa B|Transcription Factors|Human|
medline result: Antigens, CD|Antigens, Differentiation, T-Lymphocyte|B-Lymphocytes|Biological Markers|Cell Line, Tumor|DNA-Binding Proteins|Human|Lymphocyte Activation|Mutation|Protein Structure, Tertiary|Protein-Serine-Threonine Kinases|Receptors, Antigen, T-Cell|Signal Transduction|T-Lymphocytes|Transcription Factors|

from MTI to medline:
1. Protein Kinases(D011494) -> Protein-Serine-Threonine Kinases(D017346) 0.9338325
2. Receptors, Antigen, T-Cell(D011948) -> Receptors, Antigen, T-Cell(D011948) 1.0
3. Antigens, CD28(D018106) -> Receptors, Antigen, T-Cell(D011948) 0.8493386
4. Jurkat Cells(D019169) -> Cell Line, Tumor(D045744) 0.80381376
5. Protein-Tyrosine Kinase(D011505) -> Protein-Serine-Threonine Kinases(D017346) 0.84926903
6. T-Lymphocytes(D013601) -> T-Lymphocytes(D013601) 1.0
7. Antigens, CD3(D017252) -> Antigens, Differentiation, T-Lymphocyte(D000945) 0.8748187
8. Ca(2+)-Calmodulin Dependent Protein Kinase(D017871) -> Protein-Serine-Threonine Kinases(D017346) 0.8331186
9. Signal Transduction(D015398) -> Signal Transduction(D015398) 1.0
10. Mitogen-Activated Protein Kinases(D020928) -> Protein-Serine-Threonine Kinases(D017346) 0.8607602
11. Mitogen-Activated Protein Kinase Kinases(D020929) -> Protein-Serine-Threonine Kinases(D017346) 0.7586931
12. Enzyme Activation(D004789) -> Signal Transduction(D015398) 0.411908
13. Lymphocyte Specific Protein Tyrosine Kinase p56(lck)(D019860) -> Protein-Serine-Threonine Kinases(D017346) 0.624941
14. Interleukin-2(D007376) -> Biological Markers(D015415) 0.49643925
15. Receptor-CD3 Complex, Antigen, T-Cell(D017260) -> Receptors, Antigen, T-Cell(D011948) 0.7889788
16. MAP Kinase Kinase Kinases(D020930) -> Protein-Serine-Threonine Kinases(D017346) 0.75525683
17. Phosphorylation(D010766) -> Signal Transduction(D015398) 0.33803588
18. Antigens, CD2(D018801) -> Antigens, Differentiation, T-Lymphocyte(D000945) 0.7860894
19. Cell Physiology(D002468) -> Signal Transduction(D015398) 0.7366098
20. NF-kappa B(D016328) -> DNA-Binding Proteins(D004268) 0.78245896
21. Transcription Factors(D014157) -> Transcription Factors(D014157) 1.0
22. Human(D006801) -> Human(D006801) 1.0

from medline to MTI:
1. Antigens, CD(D015703) -> Antigens, CD3(D017252) 0.7605871
2. Antigens, Differentiation, T-Lymphocyte(D000945) -> Antigens, CD3(D017252) 0.8748187
3. B-Lymphocytes(D001402) -> T-Lymphocytes(D013601) 0.7273614
4. Biological Markers(D015415) -> Antigens, CD3(D017252) 0.6855663
5. Cell Line, Tumor(D045744) -> Jurkat Cells(D019169) 0.80381376
6. DNA-Binding Proteins(D004268) -> NF-kappa B(D016328) 0.78245896
7. Human(D006801) -> Human(D006801) 1.0
8. Lymphocyte Activation(D008213) -> Cell Physiology(D002468) 0.5022735
9. Mutation(D009154) -> not matched
10. Protein Structure, Tertiary(D017434) -> Signal Transduction(D015398) 0.382076
11. Protein-Serine-Threonine Kinases(D017346) -> Protein Kinases(D011494) 0.9338325
12. Receptors, Antigen, T-Cell(D011948) -> Receptors, Antigen, T-Cell(D011948) 1.0
13. Signal Transduction(D015398) -> Signal Transduction(D015398) 1.0
14. T-Lymphocytes(D013601) -> T-Lymphocytes(D013601) 1.0
15. Transcription Factors(D014157) -> Transcription Factors(D014157) 1.0

identity match:
1. Receptors, Antigen, T-Cell(D011948) -> Receptors, Antigen, T-Cell(D011948)
2. T-Lymphocytes(D013601) -> T-Lymphocytes(D013601)
3. Signal Transduction(D015398) -> Signal Transduction(D015398)
4. Transcription Factors(D014157) -> Transcription Factors(D014157)
5. Human(D006801) -> Human(D006801)

evaluation summary: 15187108|0.7820851|0.27027026

MeSH Tree Structure (from MTI to medline):

Protein Kinases(D011494) -> Protein-Serine-Threonine Kinases(D017346) 0.9338325
Protein Kinases [D08.811.913.696.620.682]
    Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]

Receptors, Antigen, T-Cell(D011948) -> Receptors, Antigen, T-Cell(D011948) 1.0

Antigens, CD28(D018106) -> Receptors, Antigen, T-Cell(D011948) 0.8493386
Receptors, Antigen, T-Cell [D12.776.543.750.705.816.824]
    Antigens, CD28 [D12.776.543.750.705.816.824.133]

Jurkat Cells(D019169) -> Cell Line, Tumor(D045744) 0.80381376
Cell Line, Tumor [A11.251.860.180]
    Jurkat Cells [A11.251.860.180.495]

Protein-Tyrosine Kinase(D011505) -> Protein-Serine-Threonine Kinases(D017346) 0.84926903
Protein Kinases [D08.811.913.696.620.682]
    Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
    Protein-Tyrosine Kinase [D08.811.913.696.620.682.725]

T-Lymphocytes(D013601) -> T-Lymphocytes(D013601) 1.0

Antigens, CD3(D017252) -> Antigens, Differentiation, T-Lymphocyte(D000945) 0.8748187
Antigens, Differentiation, T-Lymphocyte [D24.185.101.100.894]
    Antigens, CD3 [D24.185.101.100.894.095]

Ca(2+)-Calmodulin Dependent Protein Kinase(D017871) -> Protein-Serine-Threonine Kinases(D017346) 0.8331186
Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
    Ca(2+)-Calmodulin Dependent Protein Kinase [D08.811.913.696.620.682.700.125]

Signal Transduction(D015398) -> Signal Transduction(D015398) 1.0

Mitogen-Activated Protein Kinases(D020928) -> Protein-Serine-Threonine Kinases(D017346) 0.8607602
Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
    Mitogen-Activated Protein Kinases [D08.811.913.696.620.682.700.567]

Mitogen-Activated Protein Kinase Kinases(D020929) -> Protein-Serine-Threonine Kinases(D017346) 0.7586931
Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
    Mitogen-Activated Protein Kinase Kinases [D08.811.913.696.620.682.700.565]

Enzyme Activation(D004789) -> Signal Transduction(D015398) 0.411908
Biochemical Phenomena [G06.184]
    Enzyme Activation [G06.184.368]
    Signal Transduction [G06.184.850]

Lymphocyte Specific Protein Tyrosine Kinase p56(lck)(D019860) -> Protein-Serine-Threonine Kinases(D017346) 0.624941
Protein Kinases [D08.811.913.696.620.682]
    Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
    Protein-Tyrosine Kinase [D08.811.913.696.620.682.725]
        src-Family Kinases [D08.811.913.696.620.682.725.800]
            Lymphocyte Specific Protein Tyrosine Kinase p56(lck) [D08.811.913.696.620.682.725.800.315]

Interleukin-2(D007376) -> Biological Markers(D015415) 0.49643925
Biological Factors [D24.185]
    Biological Markers [D24.185.101]
    Growth Substances [D24.185.348]
        Interleukins [D24.185.348.505]

Interleukin-2 [D24.185.348.505.502]

Receptor-CD3 Complex, Antigen, T-Cell(D017260) -> Receptors, Antigen, T-Cell(D011948) 0.7889788
Receptors, Antigen, T-Cell [D12.776.543.750.705.816.824]
    Receptor-CD3 Complex, Antigen, T-Cell [D12.776.543.750.705.816.824.800]

MAP Kinase Kinase Kinases(D020930) -> Protein-Serine-Threonine Kinases(D017346) 0.75525683
Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]
    MAP Kinase Kinase Kinases [D08.811.913.696.620.682.700.559]

Phosphorylation(D010766) -> Signal Transduction(D015398) 0.33803588
Biochemical Phenomena, Metabolism, and Nutrition [G06]
    Biochemical Phenomena [G06.184]
        Signal Transduction [G06.184.850]
    Metabolism [G06.535]
        Phosphorylation [G06.535.790]

Antigens, CD2(D018801) -> Antigens, Differentiation, T-Lymphocyte(D000945) 0.7860894
Antigens, Differentiation, T-Lymphocyte [D24.185.101.100.894]
    Antigens, CD2 [D24.185.101.100.894.090]

Cell Physiology(D002468) -> Signal Transduction(D015398) 0.7366098
Cell Physiology [G04.335]
    Cell Communication [G04.335.122]
        Signal Transduction [G04.335.122.850]

NF-kappa B(D016328) -> DNA-Binding Proteins(D004268) 0.78245896
DNA-Binding Proteins [D12.776.260]
    NF-kappa B [D12.776.260.600]

Transcription Factors(D014157) -> Transcription Factors(D014157) 1.0

Human(D006801) -> Human(D006801) 1.0

MeSH Tree Structure (from medline to MTI):

Antigens, CD(D015703) -> Antigens, CD3(D017252) 0.7605871
Antigens, CD [D24.185.101.100.110]
    Antigens, CD3 [D24.185.101.100.110.095]

Antigens, Differentiation, T-Lymphocyte(D000945) -> Antigens, CD3(D017252) 0.8748187
Antigens, Differentiation, T-Lymphocyte [D24.185.101.100.894]
    Antigens, CD3 [D24.185.101.100.894.095]

B-Lymphocytes(D001402) -> T-Lymphocytes(D013601) 0.7273614
Lymphocytes [A11.118.637.555.567]
    B-Lymphocytes [A11.118.637.555.567.562]
    T-Lymphocytes [A11.118.637.555.567.569]

Biological Markers(D015415) -> Antigens, CD3(D017252) 0.6855663
Biological Markers [D24.185.101]
    Antigens, Differentiation [D24.185.101.100]
        Antigens, CD [D24.185.101.100.110]
            Antigens, CD3 [D24.185.101.100.110.095]

Cell Line, Tumor(D045744) -> Jurkat Cells(D019169) 0.80381376
Cell Line, Tumor [A11.251.860.180]
    Jurkat Cells [A11.251.860.180.495]

DNA-Binding Proteins(D004268) -> NF-kappa B(D016328) 0.78245896
DNA-Binding Proteins [D12.776.260]
    NF-kappa B [D12.776.260.600]

Human(D006801) -> Human(D006801) 1.0

Lymphocyte Activation(D008213) -> Cell Physiology(D002468) 0.5022735
Biological Phenomena, Cell Phenomena, and Immunity [G04]
    Cell Physiology [G04.335]
    Immunity [G04.610]
        Immunity, Cellular [G04.610.555]
            Lymphocyte Activation [G04.610.555.545]

Mutation(D009154) -> not matched

Protein Structure, Tertiary(D017434) -> Signal Transduction(D015398) 0.382076
Biochemical Phenomena [G06.184]
    Signal Transduction [G06.184.850]
    Molecular Conformation [G06.184.580]
        Protein Conformation [G06.184.580.709]
            Protein Structure, Tertiary [G06.184.580.709.610]

Protein-Serine-Threonine Kinases(D017346) -> Protein Kinases(D011494) 0.9338325
Protein Kinases [D08.811.913.696.620.682]
    Protein-Serine-Threonine Kinases [D08.811.913.696.620.682.700]

Receptors, Antigen, T-Cell(D011948) -> Receptors, Antigen, T-Cell(D011948) 1.0

Signal Transduction(D015398) -> Signal Transduction(D015398) 1.0

T-Lymphocytes(D013601) -> T-Lymphocytes(D013601) 1.0

Transcription Factors(D014157) -> Transcription Factors(D014157) 1.0

Appendix F: References

Aronson et al. (Oct, 1999). *The indexing Initiative*. Retrieved August 15, 2005 from http://ii.nlm.nih.gov/resources/Indexing_Initiative.pdf

Aronson et al. (n.d.). *The NLM Indexing Initiative's Medical Text Indexer*, NLM. Retrieved August 15, 2005 http://ii.nlm.nih.gov/resources/aronson-medinfo04.wheader.pdf

Bodenreider, O. (April, 2004). Semantic Similarity in Biomedical Taxonomies. *Biomedical knowledge discovery*

Indexing Initiative Project. (June, 2002). *A MEDLINE Indexing Experiment using terms Suggested by MTI*, NLM. Retrieved August 15, 2005 from http:// ii.nlm.nih.gov/resources/ResultsEvaluationReport.pdf

Kim et al. (2001) Automatic MeSH Term Assignment and Quality Assessment, NLM. *Proc AMIA Symp*. 319-23.

Lin, D. (1998) An information-theoretic definition of similarity. *In Proc. 15th International Conference on Machine Learning*

MeSH Fact Sheet (2005). Retrieved August 15, 2005 http://www.nlm.nih.gov/pubs/factsheets/mesh.html

Rada, et al (1989). Development and application of a metric on semantic nets. *IEEE Transaction on Systems, Man, and Cybernetics*, 19(1), 17-30

Resnik, P. (1996), Semantic Similarity in a Taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research.*

Semi-Automatic Indexing (n.d.) Retrieved August 15, 2005 http://ii.nlm.nih.gov/semiauto.shtml