

Qualifying Medical Web Sites

R H Baud, L E Cousineau, S Roessingh*

Medical Knowledge Discovery Project
Cousineau Consulting LLC
2806 Fennel Rd, Edgewater MD 21037
E-mail: steve.roessingh@gmail.com

*Corresponding author

Abstract: To make the enormous amount of web based health and medical information useful to users; web sites must be qualified and ranked. A methodology is investigated to identify the source and qualification of web-based health and medical information and knowledge and to rank web sites by disease class and disease based on the depth and quality of information on that disease. A detailed analysis of 3,500 web page references to 20 diseases in 3 classes showed that web sites can be identified and ranked by source class. The study developed and verified a medical page ranking methodology based on the depth of information a web site has on a disease class and disease. The resulting score is referred to as the web site Medical Information IQ. Web sites with the highest Information IQ ranking vary by disease class and disease.

Keywords: medical page ranking, Medical Information Density, Medical Information IQ, medical source category

1. Introduction

The challenge of web-based health and medical information and knowledge is the sheer quantity that is available and can be freely shared. The study addresses the issues of identifying and qualifying the origin and authenticity of web-based medical information and knowledge.

A detailed analysis of 3,500 web page references to 7 cardiovascular, 7 diabetes, and 6 skin cancer disease searches identified 9 distinct categories of medical web sites. These 9 categories of medical web sites are: general health & medical, medical publications, research articles, medical libraries, medical associations, healthcare associations, medical schools and academic, government, and healthcare & pharmaceutical industry. Within each category, the web sites can be ranked by measuring their Medical Information Density per disease class and entity. The 90 web sites with the highest Medical Information Density cover 2,100 or approximately 2/3 of the 3,500 researched web pages.

The study demonstrated that medical web sites can be assigned a “Medical Information IQ” based on the Medical Information Density per disease class that can be used to assign a page ranking and provide an assessment of the disease information quality and authentication of the web site. For research purposes, 24 searches with different information descriptors were generated for 7 related diseases in each of 3 disease classes. A total of over 5,000 web pages on 2,700 web sites were analyzed. The number of times a web site was referenced by these 24 searches was registered in the research database to measure the Medical Information Density. The web site Medical Information Density is used to calculate the “Medical Information IQ”. The study observed that there seems to be a direct relation between the “Medical Information IQ” and the quality of information displayed on web pages with the highest ranking.

The study showed that the origin and highest ranked web sites vary per disease class and disease. The medical page ranking algorithms must be applied per disease class and disease to give meaningful results.

2. Methodology

The following 7 cardiovascular, 7 diabetes, and 6 skin cancer disorders are used in the Medical Knowledge Source Study covering 20 diseases.

Cardiovascular Disorder	Diabetes Disorder	Skin Cancer
Abdominal aortic aneurysm	Alcoholic ketoacidosis	Basal cell carcinoma
Hypercholesterolemia	Blood Glucose Insulin	Carcinoma of Skin
Angina pectoris	Diabetes Mellitus	Metastasis to Skin
Hypertrophic cardiomyopathy	Diabetic cardiomyopathy	Sarcoma of Skin
Idiopathic orthostatic intolerance	Diabetic nephropathy	Skin Lymphoma
Deep Venous Thrombosis	Hypoglycemia	Skin Melanoma
Aortic Regurgitation	Hyperglycemia	

A detailed analysis is done of 3,500 knowledge web references covering the above disorders. A total of 24 searches were done per disease class and each search result filtered by 9 categories of knowledge sources. For each disorder, 3 to 4 searches are made with different context descriptors (e.g. child, female, causes, symptoms, diagnostic tests, treatments, medications, management, guidelines, outcome assessment, etc.). For each search 10 web pages are registered per source category giving a total of 90 web pages registered per search. The web site URL of each web site is recorded and an analysis is done by disease entity and class as to the number of times the web site is referenced. The number of references to a web site per disease entity and class is used to calculate a “Medical Information Density”.

3. Identifying, qualifying, and ranking web based medical information

Out of the 5,300 web pages analyzed 2,740 unique web sites were identified. The 270 web sites (10%) with the highest Medical Information Density cover over 3,200 web pages (61%).

Research Summary	Disorders	Searches	Web Pages	Web Sites	Top 90 Web Sites	% web pages
Cardiovascular Disorder	7	24	1 630	806	1 040	64%
Diabetes Disorder	7	24	1 877	948	1 102	59%
Skin Cancer	6	21	1 818	985	1 102	59%
Total Source Research	20	69	5 325	2 739	3 244	61%

3.1 Medical web site source categories

The analysis of the resulting web sites showed that they can be grouped into 9 “medical source categories” that cover all referenced web sites. The below table gives the “medical source categories” defined for research purposes.

General Health & Medical	Medical Libraries	Universities
Medical Publications	Medical Associations	Government Healthcare
Research Articles	Healthcare Organizations	Healthcare & Pharmaceutical

An analysis of the results of searches for the 20 diseases showed the following distribution of results across the source categories.

.com	.edu	.org	.gov	.nih	nlm.nih.gov	Total
41%	2%	26%	10%	5%	5%	80%

3.2 Ranking medical web pages

The over 2,700 analyzed web sites are ranked per disease class, disease, and source category according to the number of times they are referenced by disease context searches referred to as the web site Medical Information Density. This Medical Information Density is used to assign a Medical Information IQ by which web sites are ranked.

The following table gives an example of Medical Information IQ for skin cancers for Universities, Medical Associations, and General Health & Medical (.com) web sites.

Skin Cancers			Skin Cancers			Skin Cancers		
Universities			Medical Associations			General Health & Medical		
		IQ			IQ			IQ
1	umm.edu	12	1	ama-assn.org	26	1	emedicine.com	11
2	umich.edu	9	2	cancer.org	23	2	medicinenet.com	10
3	wustl.edu	9	3	aad.org	15	3	medscape.com	9
4	hartnell.edu	8	4	aacrjournals.org	14	4	healthscout.com	8
5	bcm.edu	7	5	dermnetnz.org	14	5	webmd.com	6
6	ucsf.edu	7	6	aaafp.org	12	6	about.com	5
7	virginia.edu	7	7	aacd.org	9	7	wkhealth.com	5
8	luc.edu	6	8	jco.org	8	8	essortment.com	4
9	nyu.edu	6	9	bloodjournal.org	5	9	findarticles.com	4
10	ua.edu	6	10	cap.org	5	10	muschealth.com	4

The following table compares the University web sites Medical Information IQ for cardiovascular disease, diabetes, and skin cancers. It was noted that the web site Medical Information IQ varies per disease class and disease.

Cardiovascular			Diabetes Mellitus			Skin Cancers		
Universities			Universities			Universities		
1	umich.edu	14	1	umich.edu	15	1	umm.edu	12
2	vanderbilt.edu	10	2	wustl.edu	12	2	umich.edu	9
3	nyu.edu	6	3	harvard.edu	11	3	wustl.edu	9
4	ucla.edu	6	4	ua.edu	10	4	hartnell.edu	8
5	ua.edu	5	5	virginia.edu	9	5	bcm.edu	7
6	umm.edu	5	6	nap.edu	6	6	ucsf.edu	7
7	upenn.edu	5	7	upenn.edu	6	7	virginia.edu	7
8	virginia.edu	5	8	stanford.edu	5	8	luc.edu	6
9	washington.edu	5	9	ucsf.edu	5	9	nyu.edu	6
10	jhu.edu	4	10	umm.edu	5	10	ua.edu	6

A detailed analysis of the Medical Information Density showed that it varies per disease class as illustrated by the following table that compares the highest scores per source category.

Cardiovascular Disease	Highest Ranking	Diabetes	Highest Ranking	Skin Cancer	Highest Ranking
Medical Libraries	63	Medical Libraries	77	Medical Libraries	45
Medical Associations	61	Medical Associations	75	Research Articles	37
Research Articles	41	Medical Publications	30	Government Healthcare	34
Medical Publications	33	Government Healthcare	26	Medical Publications	31
General Health & Medical	24	Research Articles	21	Medical Associations	26
Government Healthcare	17	General Health & Medical	20	Healthcare Organizations	13
Universities	14	Universities	15	Healthcare & Pharmaceutical	13
Healthcare Organizations	12	Healthcare Organizations	14	Universities	12
Healthcare & Pharmaceutical	9	Healthcare & Pharmaceutical	11	General Health & Medical	11

4. Conclusions in qualifying and ranking medical web sites

The sources of web based medical information and knowledge can be clearly identified by grouping web sites by source category. The authentication, qualifying, and ranking of web sites must be done by disease and medication class as web sites are often specialized in certain aspects of health and medical knowledge.

The greatest depth of medical information is published on the Internet by well defined categories of web sites such as medical libraries, medical associations, government health care, and universities. Medical publications rank high in depth of coverage but the articles are by paid subscription and are not open source and freely available. General health & medical web sites (.com) have less depth of information than specialized medical web sites. Within these categories, the web site rankings vary by disease class and disease.

The studies indicate a direct relationship between the depth of medical information on a disease class and the quality of the web site. This is confirmed by the high Medical Information IQ ranking of medical libraries, medical associations, and government health care.