Web Based Clinical Applications

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: Free format text descriptions of symptoms are used to identify likely associated diseases using the resources of the World Wide Web. The study made an in-depth analysis using free format text symptoms of 10 diseases in 4 disease classes. For each symptom search 50 web pages were analyzed against the Disease Ontology to identify possible associated diseases. On the average, each symptom search identified a primary disease out of 7 possible related diseases in the same disease class. The study concluded that a number of clinical information support applications can be developed using the medical resources available on the World Wide Web.

Keywords: medical symptoms, primary symptoms, secondary symptoms, primary disease, associated diseases,

1. Introduction

The web based clinical applications study investigated how free text disease attributes (e.g. symptoms, diagnostic test results) can be used to identify likely associated diseases using the resources of the World Wide Web. The research study analyzed 10 free text medical symptom searches that described symptoms in 4 disease classes; 2 cancer symptoms, 4 cardiovascular disease symptoms, 2 diabetes symptoms, and 2 musculoskeletal disease symptoms. On the average, each search identified 7 diseases with a primary disease identified by 32% of the references.

2. Methodology

For each symptom search over 50 web pages were analyzed to identify and record possible diseases. Up to 10 referenced diseases were recorded per web page and per search. The number of times it was referenced was recorded for each disease.

For each symptom search the referenced web pages were recorded in 6 source categories: as per the below table. Up to 10 web pages were recorded for each source category and each web page was analyzed for up to 10 referenced diseases and these diseases were recorded in the database. No record was maintained if the web page referenced over 10 diseases.

Medical & Scientific Journals	Universities and Medical Schools
Government Healthcare	Healthcare & Medical Organizations
Medical Libraries	General Web Sites

A summary database record was set up for each symptom search listing up to 10 diseases. For each referenced disease an entry was made how many times it was reference per source category and how many times it was referenced in total. For each symptom search up to 10 possible diseases are listed by order starting with the disease with the highest number of references.

3. <u>Medical Symptom Search</u>

The following table gives the key parameters of the symptom search research.

Number of symptom groups searched	11
Disease Classes	4
Average number of diseases identified per symptom search	7
Average number of references to diseases	32
Average number of references to the likely primary disease	13
% References that point to the likely primary disease	34%
Average number of web pages analyzed per search	53
Average number of web pages containing disease references	19

3.1 Symptoms searched

The following tables give a list per disease class of the symptoms used in the research and the intended disease for reference purposes only.

Cancer Symptoms

Cancer Symptoms	Primary Symptoms	Secondary Symptoms
Prostrate cancer	Frequent urination, weak urination, burning urination, blood in urine,	Erection difficulty, Painful ejaculation
Skin Cancer	Face skin lesion brown,	scar sore bleeding

Cardiovascular Symptoms

Cardiovascular Symptoms	Primary Symptoms	Secondary Symptoms
Abdominal aortic aneurysm	Intense pain lumbosacral region, abdominal pulsation,	Back pain
Aortic Regurgitation	Systolic depression left parasternal area, low aortic diastolic pressure,	Fever, anemia, weight loss, embolic phenomena
Atrial Filibration	Irregular pulse, palpitations, chest restriction,	Light-headedness, dizzyness, dyspnea
Deep Venous Thrombosis	Swelling of the lower left calf, swollen veins,	Sensitivity and pain in the lower left leg

Diabetes Symptoms

Diabetes Symptoms	Primary Symptoms	Secondary Symptoms
Autonomic neuropathy	Orthostatic hypotension, dysphagia, exercise intolerance,	Nausea, diarrhea, urinary retention
Diabetic retinopathy	Focal blurring, retinal detachment, partial visual loss,	

Musculoskeletal Symptoms

Musculoskeletal Symptoms	Primary Symptoms	Secondary Symptoms
Gout	Sudden toe pain, swollen toe, severe toe pain,	White papules toe joint
Sjögren's syndrome	Dry Eye, damaged cornea, impaired vision,	Diminished saliva, dry throat

3.2 Symptom search study results

The following tables give a summary of each symptom search by order of the % of web page references to the primary disease. The following gives the terminology and interpretation of the results in the tables.

- **Symptom Search** gives the name of the target disease for research and identification purposes only.

- **Primary disease** gives the identified disease with the highest number of web page references.

- **% references** give the % of all disease references that identify the primary disease.

- **Diseases referenced** give the number of diseases identified on the analyzed web pages.

- **Analyzed pages** gives the number of web pages analyzed for the symptom search

- **Pages with diseases** refer to the number of analyzed web pages that were found to contain recognizable references to diseases. Web pages with references to over 10 diseases were eliminated.

Symptom Search	Primary Disease	% of references	Diseases Referenced	Disease References	Analyzed Pages	Pages with diseases
Deep Venous Thrombosis	Deep venous thrombosis	71%	10	56	59	42
Prostrate cancer	Prostrate cancer	66%	7	51	50	33
Skin Cancer	Squamous cell carcinoma	50%	8	75	60	30
Sjögren's syndrome	Sjogren's syndrome	36%	9	72	53	19
Mixed connective tissue disease	rheumatoid arthritis	30%	7	21	44	13
Diabetic retinopathy	Age related macular degeneration	27%	10	26	60	16
Atrial Filibration	myocardial infarctions	25%	10	18	53	13
Autonomic neuropathy	Diabetic Neuropathy	21%	5	11	53	11
Abdominal aortic aneurysm	Abdominal Aortic Aneurysm	20%	7	10	54	11
Aortic Regurgitation	Myocardial infarction	18%	8	9	40	7
Gout	lupus erythematosus	2%	1	1	55	1

4. <u>**Symptom search study conclusions**</u>

The research on 11 symptom searches provided very varying results with only 3 of the searches identifying the primary disease with over 50% of the analyzed disease references. The number of diseases corresponding to the symptom search varied from 1 to 10 diseases with an average of 7.4 diseases identified per search. In the majority of the searches, the identified diseases were closely related and in the same disease class.

The study found that a major factor in the capability to identify likely diseases and the primary disease is the number of web pages in the search results that contain identifiable disease references. On the average, 53 web pages were analyzed per symptom search, and of these an average of 19 web pages contained identifiable disease references. The number of pages containing identifiable disease references ranged from 1 for the gout symptoms to 42 for the deep venous thrombosis symptoms. The number of pages with identifiable diseases directly affects the ability to clearly identify a primary disease.

There seems to be a direct relationship between the ability to clearly identify a primary disease and the number of disease references in the analyzed web pages. The number of disease references in the analyzed web pages varied from 1 (Gout) to 75 (Skin cancer). For the symptom searches where less then 30 disease references could be identified less then 30% of the references identified a primary disease.

Two types of web pages references by Google in the search results when symptoms are entered make it very difficult to analyze the web page and determine diseases corresponding to the described symptoms.

- Web pages containing long lists of diseases with corresponding symptoms.
- Web pages containing several case studies of patients with diverging symptoms and disorders.

The objective of further research is to define methods by which the proximity of the symptoms to disease terms is used to identify associated diseases. The study did not take into account the number of times a disease is referenced within one web page. Developing methodologies, structures and algorithms to take this into account would also assist in determining the primary disease associated with the symptoms.

With further research, web based applications can be developed to identify with reasonably accuracy the primary and associated diseases associated with symptom descriptions.