Medical Context Descriptors

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

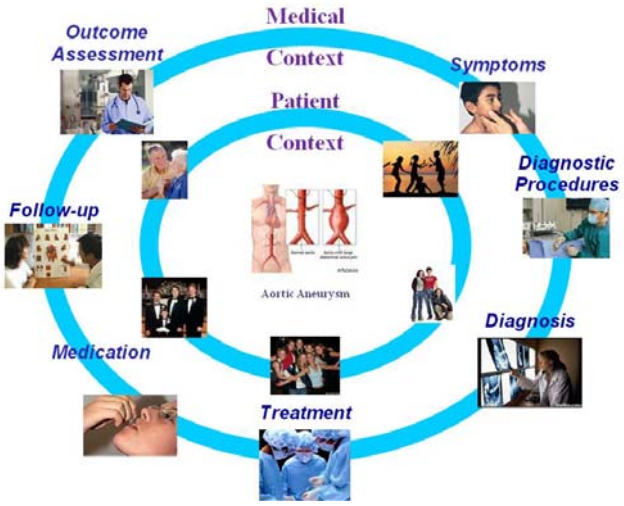
Abstract: The study investigates how medical context driven web based health and medical information and knowledge can be identified. The study identified and tested over 500 combinations of disease and medication context descriptors. A methodology is proposed to formally define and apply context descriptors to disease and medication classes. The structure for context driven health and medical knowledge applications is proposed as well as a methodology to test and validate context descriptors against medical ontologies.

Keywords: disease context descriptors, medication context descriptors, patient context, medical process context

1. Introduction

The context and content of health and medical information and knowledge must be associated to provide meaningful web based knowledge systems to patients, physicians, health care professionals, academics, and students. Medical Context Descriptors are terms most often found on health and medical web pages that describe the context of the information and knowledge on diseases, disorders, medications and drugs. The study investigates how Medical Context Descriptors can be defined and mapped to diseases and medications.

Descriptors refer to the information context about an ontology entity and not the entity itself. For example the combination of “overweight child angina pectoris outcome assessment” does not describe angina pectoris but the information context about angina pectoris.



2. Methodology

A medical information descriptor matrix of 500 combinations was created and investigated against 10 diseases in 4 disease classes as per the below table. 200 combinations were applied to all 10 diseases (2,000 searches) and all 500 combinations were applied to 2 diseases (angina pectoris, basal cell carcinoma). The number of web page references was recorded for each combination of medical information descriptors.

Diseases Researched

Cardiolovascular Disorders
Angina Pectoris
Coronary artery spasm
Skin Cancer
Basal cell carcinoma
Skin Lymphangiosarcoma
Diabetes mellitus
Diabetic nephropathy
Glomerular sclerosis
Hematology & Oncology Disorders
Iron deficiency anemia
Myelofibrosis
Musculoskeletal Disorders
Sjögren's syndrome
Tendinopathy

3. Medical Concept Descriptors

The study analyzed web based information for over 700 cardiovascular, diabetes, hematology and oncology, and musculoskeletal diseases and how the medical context is described for these diseases.

The medical context descriptors used for the study are defined by the World Health Organization, WHO and recognized clinical terminology.

Morphology
(Morphology descriptor)

Sex / Age Group
(Patient reference descriptors)

Disease, Disorder, or Sign
(Medical Term)

Medical Practice / Medical Background / Medical Process
(Patient process descriptors)

3.1 Morphology

The patient morphology defined by the World Health Organization is used throughout all medical literature and medical references with some minor variations. They are based on the Body Mass Index BMI standards and definitions set by the WHO. These are:

- **Overweight** - BMI 25,0-29,9 = overweight
- **Obese** - BMI 30,0-34,9 = obese class 1
- **Severely Obese** - BMI 35,0 and over = obese class 2 / 3

The study showed these descriptors combined with patient and patient process descriptors apply to almost all knowledge references to specific diseases and disorders.

3.2 Patient

The patient descriptors used by the World Health Organization are reflected in most medical literature and references. Some variations are used in non-scientific publications and references, but they do not reflect the main stream of medical literature.

Patient descriptors defined by WHO and most commonly used are as follows.

- **Child**- A person between birth and puberty
- **Adolescent**- The development of children ages 12 through 18 years old
- **Female**- Adult female "an individual of the sex that bears young" and female chromosome constitution.
- **Male**- Adult male "an individual of the sex that produces sperm" and male chromosome constitution.
- **Elderly** – “Most developed world countries have accepted the chronological age of 65 years as a definition of 'elderly' or older person” – WHO.

3.3 Patient Process

Patient process descriptors are standardized in Evidence Based Medicine EBM practice and related medical literature and references. Patient process descriptors can be described in 3 broad categories.

- **Medical practice** descriptors that describe the case-disease process of the patient as followed in EBM practice.
- **Medical background** descriptors that describe the medical background of the patient and the case-disease.
- **Medical process** descriptors that describe the various medical processes that can affect the case-disease of the patient

3.4 Medical context descriptor classes

For purposes of the research, 5 medical context descriptor classes were defined: morphology, patient, medical practice, medical background, and medical process. For the research, it was defined that the morphology and patient descriptors can be combined (25 combinations) with the 3 medical process categories giving a total of 500 context descriptor combinations per disease as illustrated in the below table.

Descriptors	**# of Descriptors**	**Combinations**		**Total Combinations**
Combinations				**500**
Morphology	4			
Patient	6	((6*4)+1)	25	
Medical Practice	8	24+1	25	200
Medical Background	5	24+1	25	125
Medical Process	7	24+1	25	175

4. Applying medical context descriptors

A number of analysis were done on the effect of applying disease context descriptors and the ability to refine the search for web based medical information.

4.1 The effect of medical context descriptors

The following table summarizes the average number of web page references per disease and per descriptor class.

Diseases Researched	Web Pages	Medical Practice	Medical Background	Medical Process	Medical Average	Weight Average	Patient Average
Angina Pectoris	1 870 000	359 114	834 400	351 000	514 838	61 249	139 861
		19,2%	44,6%	18,8%	27,5%	3,3%	7,5%
Basal cell carcinoma	3 250 000	601 929	628 540	712 100	647 523	76 649	271 551
		18,5%	19,3%	21,9%	19,9%	2,4%	8,4%
Coronary artery spasm	165 000	72 586	79 920	49 129	67 211	19 299	44 845
		44,0%	48,4%	29,8%	40,7%	11,7%	27,2%
Diabetic nephropathy	2110000	227 129	389 800	300 614	305 848	116 501	148 153
		10,8%	18,5%	14,2%	14,5%	5,5%	7,0%
Glomerular sclerosis	103 000	29 666	43 400	49 893	40 986	10 699	21 094
		28,8%	42,1%	48,4%	39,8%	10,4%	20,5%
Iron deficiency anemia	839 000	541 286	267 680	480 571	429 846	177 251	301 958
		64,5%	31,9%	57,3%	51,2%	21,1%	36,0%
Myelofibrosis	318 000	63 909	70 720	71 103	68 577	8 318	27 626
		20,1%	22,2%	22,4%	21,6%	2,6%	8,7%
Sjögren's syndrome	945 000	508 029	330 800	459 843	432 890	44 309	89 195
		53,8%	35,0%	48,7%	45,8%	4,7%	9,4%
Skin Lymphangiosarcoma	13 700	7 917	5 816	9 423	7 719	1 185	2 943
		57,8%	42,5%	68,8%	56,3%	8,6%	21,5%
Tendinopathy	93 100	24 299	32 582	12 404	23 095	3 489	8 826
		26,1%	35,0%	13,3%	24,8%	3,7%	9,5%

The research shows that applying medical context descriptors to diseases produces a high number of web page references but does assist in narrowing down the relevant web pages by 50% to 92% depending on the descriptor combination used. The study showed that the majority of the web pages contain information relevant to the applied information descriptors so that the user can specify to a good degree of accuracy what information he is seeking about a disease.

Applying the above research results to the average of 1 Million web pages per disease, the following table illustrates the effect of applying medical context descriptors.

Average Disease	Medical Practice	Medical Background	Medical Process	Medical Average	Weight Average	Patient Average
1 000 000	343 542	339 604	343 545	342 231	74 064	155 591
	34%	34%	34%	34%	7%	16%

An analysis was done of the effect of each medical context descriptor in the 5 descriptor classes and the % of web pages referenced when applied to a disease entity. The research analysis indicated that medical context descriptors with the highest % web page references seem to be the most relevant in defining disease information searches.

Morphology / Weight descriptors

Obese seems to be the most relevant weight/ morpholohy descriptor applied to disease entities to find relevant web pages.

Overweight	Obese	Severely Obese	Weight Average
3%	5%	1%	3%

Patient descriptors

The descriptors “child”, “female”, and “pregnancy” are important factors in searching for disease information and knowledge.

Child	Adolescent	Female	Pregnancy	Male	Elderly	Patient Average
42,8%	7,1%	44,3%	43,3%	25,4%	12,5%	29,3%

Medical Practice descriptors

Diagnostic	Diagnosis	Treatment	Medication	Follow-up	Tracking	Outcome Assessment	Practice Average
30,2%	44,9%	53,9%	44,0%	19,1%	5,5%	32,9%	34,4%

Medical Background descriptors

Symptoms	Causes	Risk factors	History	Family History	Background Average
40,4%	43,6%	29,6%	28,9%	17,3%	34,0%

Medical Process descriptors

Definition	Guidelines	Best practices	Compliance	Alternative medicine	Clinical trials	Genes	Process Average
65,3%	64,3%	8,1%	7,2%	13,8%	19,5%	30,7%	34,4%

4.2 Applying medical context descriptors

The following are examples of the results of adding disease context descriptors relevant to the case – disease profile of the patient. Medical search applications can be developed that use interactive menus to add information descriptors to disease entities.

The research illustrates that even complex medical information requests generate links to tens of thousands web page references. Therefore, an effective method to categorize, prioritize, and filter reference sources is essential.

Cardiolovascular Disorders	**Reference Request**	**Disorder References**	**Request References**
Abdominal aortic aneurysm	Adolescent abdominal aortic aneurysm guidelines	1'000'000	43'000
Coronary artery spasm	Overweight adolescent Coronary artery spasm causes	440'000	12'000
Orthostatic intolerance	Child orthostatic intolerance medication	198'000	200'000
Deep Venous Thrombosis	Obese Elderly Deep Venous Thrombosis follow-up	1'200'000	83'000

Diabetes mellitus	**Reference Request**	**Disorder References**	**Request References**
Diabetic nephropathy	Overweight child diabetic nephropathy outcome assessment	1'180'000	56'000
Glomerular sclerosis	Elderly glomerular sclerosis diagnostic tests	420'000	51'000
Hypokalemia	Obese female hypokalemia medication	690'000	30'000
Urine glucose	Child urine glucose tracking	1'300'000	130'000
Physiologic hypoglycemia	Overweight male physiologic hypoglycemia symptoms	200'000	30'000

Musculoskeletal Disorders			
Gonococcal arthritis	Child Gonococcal arthritis symptoms	120'000	51'000
Polymyalgia rheumatica	Overweight male polymyalgia rheumatica diagnostic tests	400'000	10'000
Raynaud's phenomenon	Obese female raynaud's phenomenon diagnosis	400'000	15'000
Rotator cuff tendinitis	Adolescent rotator cuff tendinitis treatment	280'000	15'000
Hypomagnesemia	Elderly Hypomagnesemia medication	250'000	57'000
Secondary osteoporosis	Female Secondary osteoporosis outcome assessment	1'150'000	310'000

Rare Disorders	**Reference Request**	**Disorder References**	**Request References**
Musculoskeletal Disorders			
Calcific periarthritis	Obese Elderly Calcific periarthritis risk factors	700	122
Jaccoud's arthritis	Overweight Child Jaccoud's arthritis guidelines	500	4
Adventitial bursitis	Obese Elderly Adventitial bursitis outcome assessment	700	60
Epiphysitis of the calcaneus	Female Epiphysitis of the calcaneus medication	450	120
Chondromyxofibroma	Overweight Child Chondromyxofibroma treatment	500	1

4.3 Search Engine effectiveness comparison

The study investigated Internet Search Engines in their ability to distinguish complex combinations of medical context descriptors with disease terms. Six Internet Search Engines were compared against 12 complex medical search requests. The first 10 links were analyzed for relevant content. Internet Search Engines tested were: Google, Google Scholar, Yahoo, MSN Search, SearchMedica, MedStory. The maximum score is 120.

Intelligent Search Term	Google	Google Scholar	Yahoo	MSN	Search Medica	Med Story
Overweight bipolar disorder medication	9	10	10	10	7	8
Female bipolar disorder diagnosis	8	10	10	10	10	8
Obese child chronic fatique syndrome	10	10	5	2	5	6
Adolescent acne history	10	10	10	10	10	9
Obese female acne diagnostic tests	9	9	7	8	10	6
Child Blood Glucose Insulin treatment	9	9	10	9	10	8
Adolescent abdominal aortic aneurysm guidelines	2	3	5	7	10	8
Overweight child diabetic nephropathy outcome assessment	9	9	4	3	0	3
Overweight child diabetic nephropathy medication	9	10	9	8	10	6
Child Gonococcal arthritis symptoms	8	10	8	10	10	10
Obese female Raynaud's phenomenon diagnosis	8	10	2	6	9	6
Overweight Child Singulair Recommended Dose	8	7	8	7	9	4
Total "Search Score"	**99**	**107**	**88**	**90**	**100**	**82**

The below table gives the "Search Score" for the 6 search engines that are tested with the above medical context descriptors.

Search Engine	Search Score
Google Scholar	107
Search Medica	100
Google	99
MSN	90
Yahoo	88
Med Story	82

5. Medical Context Descriptor study conclusions

The study showed that medical context descriptors can be identified and formally defined per disease and medication class to identify context driven web based medical information and knowledge. The study showed that the context of web based health and medical information for a large majority conforms to internationally recognized and defined terminology.

The study identified a formal methodology to validate medical context descriptors and their combinations per disease and medication.

The Internet Search Engine benchmark tests showed that the most commonly used Internet Search Engines differentiate between medical terms and medical context descriptors.