

Semantic Components: A model for enhancing retrieval of domain-specific information

Lecture 22
CS 410/510
Information Retrieval on the Internet

Pathway Project Team

- Susan Price, MD Portland State University
- Lois Delcambre, PhD, Portland State University
- Marianne Lykke Nielsen, PhD, Royal School of Library & Information Science
- Tim Tolle, PhD, Hydrology USDA Forest Service, retired
- Vibeke Luk, MLS sundhed.dk
- Mat Weaver, PhD, CS EarthSoft, Inc.

Supporting domain experts using domain-specific digital libraries

- Domain experts often have specific information needs, perhaps related to a particular task
 - Example: physician seeing pt with chronic asthma, newly pregnant
 - Retrieved documents should be relevant to the task or question (not just about the topic)
 - Time for searching may be limited

CS 510 Winter 2007

3

Focus

- Support users who
 - Know how domain information is organized and expressed
 - Know types of documents available
- Using domain-specific digital libraries
 - Documents + retrieval system
 - Often lacks benefit of web hyperlink structure
- For precision-oriented information needs
 - Specific need
 - Satisfied by one, or a few, documents

CS 510 Winter 2007

4

Excerpted and condensed from http://www.emedicinehealth.com/asthma/article_em.htm

ASTHMA

Disease or Condition

Focus

Asthma Overview

Asthma is a disease that affects the breathing passages of the lungs (bronchioles). Asthma is caused by chronic (ongoing, long-term) inflammation of these passages. When the inflammation is "triggered" by any number of external and internal factors, the passages swell and fill with mucus. Muscles within the breathing passages contract (bronchospasm), causing even further narrowing of the airways. This narrowing makes it difficult for air to be breathed out (exhaled) from the lungs. This resistance to exhaling leads to the typical symptoms of an asthma attack.

Like any other chronic disease, asthma is a condition you live with every day of your life. You can have an attack any time you are exposed to one of your triggers. Asthma cannot be cured, but it can be controlled. With proper treatment, people with asthma can have fewer and less severe attacks. Without treatment, they frequent and more severe asthma attacks and can even die.

Asthma is a very common disease in the United States, where more than 17 million people are affected. A third of these are children. In 2002, 478,000 hospitalizations and 4,657 deaths were attributed to asthma.

- Asthma affects all ages and is slightly more common in African Americans than in other races.
- Asthma affects all races, although it is more common in younger people. The frequency and severity of asthma attacks tend to decrease as a person ages.

Asthma Causes

We do not know exactly what causes asthma. What all people with asthma have in common is chronic airway inflammation and excessive airway sensitivity to various triggers ... Each person with asthma has his or her own unique set of triggers. Most triggers cause attacks in some people with asthma and not in others. Common triggers of asthma attacks are the following:

- Exposure to tobacco or wood smoke
- Breathing polluted air
- Inhaling other respiratory irritants such as perfumes or cleaning products
- Breathing in allergy-causing substances (allergens) such as molds, dust, or animal dander
- An upper respiratory infection, such as a cold, flu, sinusitis, or bronchitis

Risk factors for developing asthma

- Hay fever (allergic rhinitis) and other allergies - The single biggest risk factor
- Eczema - Another type of allergy affecting the skin
- Genetic predisposition - A parent, brother, or sister also has asthma

Asthma Symptoms

When the breathing passages become irritated or infected, an attack is triggered. The attack may come on suddenly or slowly over several days or hours. The main symptoms that signal an attack are as follows: Wheezing; Breathlessness; Chest tightness; Coughing

Exams and Tests

Measurements of how well you are breathing include the following:

- Spirometer:** This device measures how much air you can exhale and how forcefully you can breathe out. The test may be done before and after you take inhaled medication. Spirometry is a good way to see how much your breathing is impaired during an attack.
- Peak flow meter:** This is another way of measuring how forcefully you can breathe out during an attack.
- Pulse oximetry:** A painless probe, called a pulse oximeter, will be placed on your fingertip to measure the amount of oxygen in your bloodstream.

A **chest x-ray** may also be taken. This is mostly to rule out other conditions that can cause similar symptoms.

Asthma Treatment

If you are in the emergency room, treatment will be started while the evaluation is still going on.

- You may be given oxygen through a face mask or a tube that goes in your nose.
- You may be given aerosolized beta-agonist medications through a face mask or a nebulizer, with or without an anticholinergic agent.
- Another method of providing inhaled beta-agonists is by using a **spacer** device. An MDI with a spacer delivers a standard dose of medication per puff. MDIs are often used along with a "spacer" or holding chamber. A dose of 2 puffs is sprayed into the spacer, which is then inhaled. The advantage of an MDI with a spacer is that it requires little or no assistance from the respiratory therapist.

Etiology

Prognosis

Epidemiology

Etiology

Epidemiology

Diagnosis

Treatment

5

Basic idea: search on main topic; refine by searching within components

find information about treating asthma with salmeterol during pregnancy

Document about asthma

Behandling

treatment

specific drug

pregnancy

Publiseret 1. oktober 2005
Seneste faglige revidering 30. januar 2005
Seneste redaktionelle revidering 30. januar 2005
Artikel-id: 1023050811124512
Titel: [Titel](#)
Dato:

Our approach

- Leverage user's knowledge
 - of the types of documents available
 - of the kinds of information in the documents
 - by allowing user to specify search using domain-specific components of documents (not necessarily structural)

CS 510 Winter 2007

7

Our approach

- Supplemental indexing that allows search within segments of documents
- Orthogonal to (and can be combined with) other indexing techniques
 - Full text indexing
 - Keyword indexing
 - Subject description
 - Other metadata

CS 510 Winter 2007

8

Our settings

- sundhed.dk: national Danish health portal
 - Serves needs of clinicians and citizens
 - 24,000+ documents
 - In use since 2001
- USDA Forest service
 - Primarily documents mandated by the National Environmental Protection Act (NEPA)
 - Research portal

CS 510 Winter 2007

9

Semantic component model

- Document classes (genres)
 - Classifications of documents: type of topic, purpose
 - Documents: about a *disease*, about a *clinical method*, about a *drug*, about a *clinical unit*
- Semantic components
 - Each document class associated with a small set of semantic components
 - Document about a disease: *treatment, evaluation, referral*
 - Document about a drug: *target group, side effects, indications*
- Semantic component instances
 - Segments of text with information about a semantic component
 - Variable length, may be nested or discontinuous

CS 510 Winter 2007

10

ERCP - Endoskopisk Retrograd Cholangio Pancreaticografi

Ved hjælp af en mavekikkert (et gastroskop) fremstilles billeder af bugspytkirtlen og galdevejene.

Hvor foregår undersøgelsen?
Undersøgelsen foregår på røntgenafdelingen. *practical*

Hvad er en ERCP?
Til ERCP bruges en tynd bæjelig slange, hvorigennem man kan se direkte ind på slimhinden i tolvfingertarmen. Lægen kan her lokalisere udførselsgangen fra bugspytkirtlen og galdevejene. Ved hjælp af en tynd sonde, som føres ned gennem skopet, kan der sprøjtes kontrastvæske i udførselsgangen, så bugspytkirtlen og galdevejene kan ses på røntgenbilleder. *description*

Forberedelse til undersøgelsen
Inden undersøgelsen skal der foreligge nye blodprøvesvar; disse tages efter aftale med sygeplejersken.
For at undersøgelsesresultatet bliver bedst muligt, bedes du faste fra kl. 24.00 aftenen før (dvs. ikke spise, drikke eller ryge, og ingen medicin). *preparation*
Du vil inden undersøgelsen få mulighed for at tale med den sygeplejerske, som er med ved undersøgelsen – og kan dér stille evt. spørgsmål.

Medicin
Hvis du får medicin, skal du medbringe morgenmedicin. *practical*

Selve undersøgelsen
Når du møder på røntgenafdelingen, vil du få en seng og få udleveret hospitalstøj). *description*
Du bliver kært til på første sal, hvor undersøgelsen vil foregå. Inden undersøgelsen bliver du lokalbedøvet i svælget med en spray og gennem en kanyler i armen får du afslappende og smertestillende medicin. Under undersøgelsen skal du ligge på et røntgenleje i Natostilling (mave/sideleje)
Gastroskoptet føres gennem munden ned i spiserøret og mavesækken til tolvfingertarmen, hvorefter der indsprøjtes en lille mængde røntgenkontrast i bugspytkirtlens og galdegangens udførselsgang.
Derefter foretages en røntgenundersøgelse af bugspytkirtlen og galdeveje.
Der pustes luft ned i maven under undersøgelsen; dette kan give ubehag i form af trykkende fornemmelse i maven og evt. opstød.
Ligeledes kan indsprøjtning af kontrasten i udførselsgangene give ubehag.
Undersøgelsen varer ca. 1 time.

Efter undersøgelsen
Du bliver kært tilbage til afdelingen. På grund af den afslappende medicin vil du være noget træt, måske sovende, efter undersøgelsen. Der skal medbringes på 1 time efter undersøgelsen, får du må spise og drikke.
Under normale omstændigheder kan du forlade afdelingen *senest på dagen*.
Du må ikke køre bil resten af døgnet pga. den afslappende medicin. *Risk*

Praktiske oplysninger
Da undersøgelsestiden varierer meget, kan der let opstå ventetid.
Du bedes være indstillet på ventetiden og evt. medbringe fx læsestof.
Det anbefales, at du ikke medbringer værdigenstande, smykker, penge o.lign., da disse ved bortkomst ikke erstattes af sygehuset. *practical*

Mere information
Kontakt redaktionen

11

Using semantic components

- Searching for documents with particular semantic components
 - Allow user to specify aspects of interest
- Searching within semantic components
 - Focus search on terms associated with a particular aspect of a topic
- Profiling documents in search results
 - Help user decide which documents to look at

Four main areas of inquiry

1. Are semantic components useful for retrieving documents?
2. How easily can semantic components be identified and represented in an index?
3. Can searchers express information needs using document types and semantic components?
4. Can document classes and semantic components be identified for a particular domain-specific document collection?

CS 510 Winter 2007

13

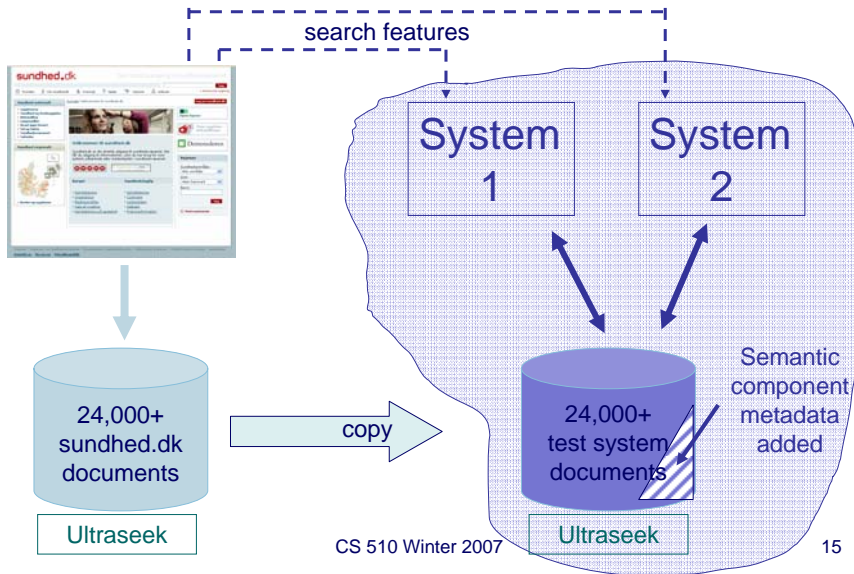
1. Searching with semantic components

- Can physicians using a search system with semantic components successfully complete search scenarios **more quickly** than using a basic search system without semantic components?
- Do physicians' searches using a search system with semantic components result in **better search performance** than when using a basic system without semantic components?

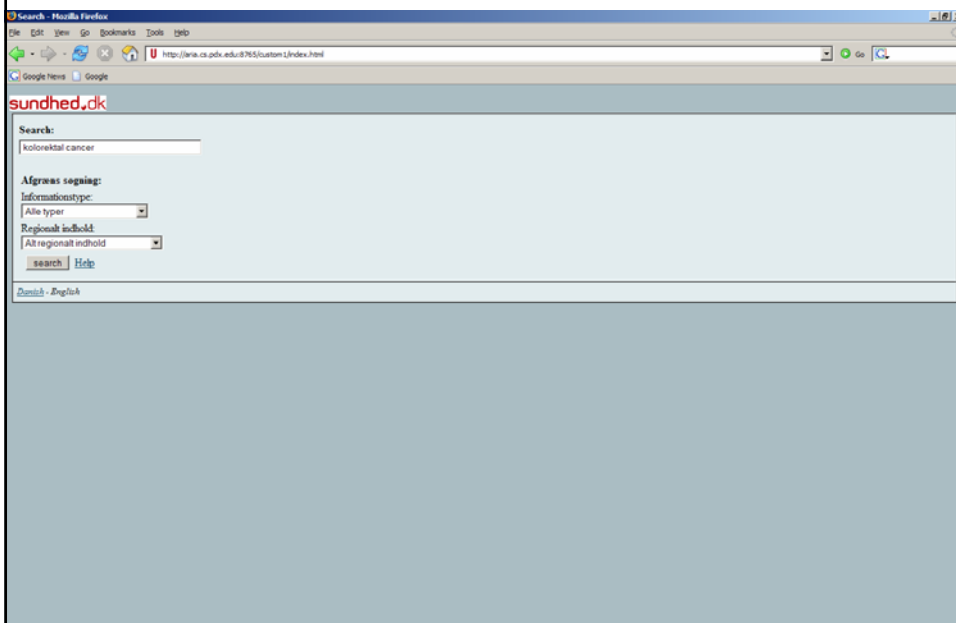
CS 510 Winter 2007

14

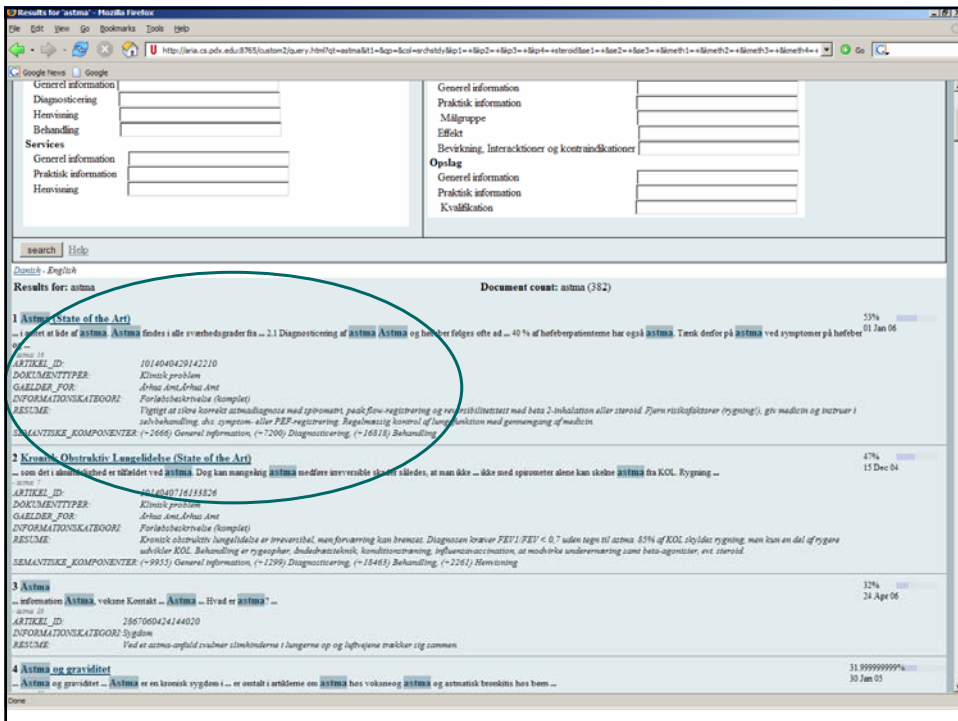
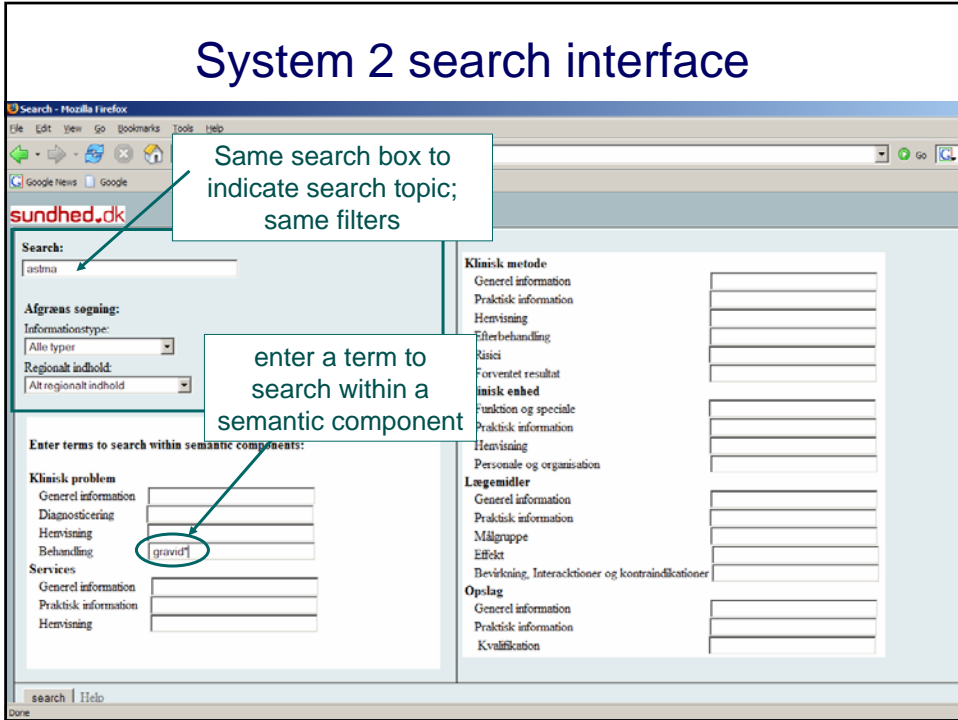
Experimental search system



System 1 ("Basic") interface



System 2 search interface



Search study experimental design

- Subjects: 30 Danish physicians who use sundhed.dk
- Intro/training to each system with guided practice
- Tasks: four search scenarios
 - Realistic, representative tasks
 - Designed in collaboration with a Danish physician/researcher
 - Two tasks with each system (randomized sequence of tasks and systems)

CS 510 Winter 2007

19

Search study experimental design (2)

- Data collection:
 - Search logs: queries, time, search results
 - Graded relevance assessments (explicit user relevance, implicit via click-through)
 - User satisfaction and feedback (questionnaires)

CS 510 Winter 2007

20

Searching study: Preliminary results

1. Do semantic components improve speed?

Clock time

System	Average duration per session				
	Scenario A	Scenario B	Scenario C	Scenario D	All scenarios
1	04:26	08:49	04:53	05:03	05:48
2	06:05	09:15	07:35	05:47	07:10

Search iterations

System	Number of search iterations per session (mean \pm std. dev.)				
	Scenario A	Scenario B	Scenario C	Scenario D	All scenarios
1	1.67 \pm 1.18	3.93 \pm 2.15	1.87 \pm 1.36	2.67 \pm 1.88	2.53 \pm 1.87
2	1.60 \pm 0.99	4.33 \pm 2.72	3.47 \pm 2.59	3.33 \pm 2.23	3.18 \pm 2.40

CS 510 Winter 2007

21

Searching study: Preliminary results

2. Do sem. comp. improve search performance?

User perspective

- Number of successful search sessions
- Rank of “best” relevant document

System perspective

- Ranking of all relevant documents
 - Based on reference standard
 - “Best query” versus session discounting
 - Using MAP
 - Using graded relevance metric

CS 510 Winter 2007

22

Searching study: User perspective

- Number of successful search sessions

System	Number of successful sessions				
	Scenario A	Scenario B	Scenario C	Scenario D	All scenarios
1	14	9	15	14	52
2	15	10	15	15	55

- Rank of “best” relevant document

System	Average reciprocal rank of the best user-relevant document				
	Scenario A	Scenario B	Scenario C	Scenario D	All scenarios
1	0.75	0.30	0.45	0.53	0.51
2	0.83	0.47	0.38	0.58	0.57

CS 510 Winter 2007

23

Searching study: System perspective

“Best” query in each search session

- Mean average precision (highly relevant docs)

System	Mean Average Precision				
	Scenario A	Scenario B	Scenario C	Scenario D	All scenarios
1	0.28	0.53	0.21	0.19	0.31
2	0.56	0.58	0.26	0.27	0.42

- Discounted cumulative gain

System	Average Normalized Discounted Cumulative Gain				
	Scenario A	Scenario B	Scenario C	Scenario D	All scenarios
1	0.41	0.59	0.37	0.34	0.43
2	0.60	0.60	0.48	0.46	0.53

CS 510 Winter 2007

24

2. Indexing with semantic components

- Is semantic component indexing of sundhed.dk documents **more consistent** than keyword indexing of the same documents?
- Is semantic component indexing of sundhed.dk documents **more accurate** than keyword indexing compared to a reference standard?
- Is semantic component indexing of sundhed.dk documents **faster** than keyword indexing?
- Is semantic component indexing of sundhed.dk documents **easier** than keyword indexing, as perceived by the indexers?

Indexing study experimental design

- Subjects: 16 Danish indexers
 - who keyword index documents for sundhed.dk
- Tasks: 12 existing sundhed.dk documents
 - Index 6 documents with semantic components
 - Index 6 documents with keywords
- Data collection:
 - Indexing data (on paper to avoid UI issues)
 - Time
 - User ease, confidence, satisfaction, and feedback (questionnaires)

CS 510 Winter 2007

27

Semantic component indexing

ERCP - Endoskopisk Retrograd Cholangio Pancreaticografi

Ved hjælp af en mavekikkert (et gastroskop) fremstilles billeder af bugspytkirtlen og galdevejene.

Hvor foregår undersøgelsen?
Undersøgelsen foregår på røntgenafdelingen. *practical*

Hvad er en ERCP?
Til ERCP bruges en tynd bæjelig slange, hvorigennem man kan se direkte ind på slimhinden i tolvfingertarmen. Lægen kan her lokalisere udførselsgangen fra bugspytkirtlen og galdevejene. Ved hjælp af en tynd sonde, som føres ned gennem skopet, kan der sprøjtes kontrastvæske i udførselsgangen, så bugspytkirtlen og galdevejene kan ses på røntgenbilleder. *description*

Forberedelse til undersøgelsen
Inden undersøgelsen skal der foreligge nye blodprøvesvar; disse tages efter aftale med sygeplejersken. For at undersøgelsesresultatet bliver bedst muligt, bedes du faste fra kl. 24.00 aftenen før (dvs. ikke spise, drikke eller ryge, og ingen medicin). *preparation*

Du vil inden undersøgelsen få mulighed for at tale med den sygeplejerske, som er med ved undersøgelsen – og kan der stille evt. spørgsmål.

Medicin
Hvis du får medicin, skal du medbringe morgenmedicin. *practical*

Selve undersøgelsen
Når du møder på røntgenafdelingen, vil du få en seng og få udleveret hospitalstøj. Du bliver kørt til på ferise sal, hvor undersøgelsen vil foregå. Inden undersøgelsen bliver du lokalbedøvet i svælget med en spray og gennem en kanyte i armen får du afslappende og smertestillende medicin. Under undersøgelsen skal du ligge på et røntgenleje i Natostilling (mave/sideleje). *description*

Gastroskoptet føres gennem munden ned i spiserøret og mavesækken til tolvfingertarmen, hvorefter der indsprøjtes en lille mængde røntgenkontrast i bugspytkirtlens og galdegangens udførselsgang. Derefter foretages en røntgenundersøgelse af bugspytkirtlen og galdeveje. Der pustes luft ned i maven under undersøgelsen; dette kan give ubehag i form af trykkende fornemmelse i maven og evt. opstød. Ligeledes kan indsprøjtning af kontrasten i udførselsgangene give ubehag. Undersøgelsen varer ca. 1 time.

Efter undersøgelsen
Du bliver kørt tilbage til afdelingen. På grund af den afslappende medicin vil du være noget træt, måske sovende, efter undersøgelsen. Der skal mindst gå 1 time efter undersøgelsen, før du må spise og drikke. Under normale omstændigheder kan du forlade afdelingen senere på dagen. *risik*

Du må ikke køre bil rosten af døgnet pga. den afslappende medicin.

Praktiske oplysninger
Da undersøgelsestiden varierer meget, kan der let opstå ventetid. Du bedes være indstillet på ventetiden og evt. medbringe fx læsestof. Det anbefales, at du ikke medbringer værdigenstande, smykker, penge o.lign., da disse ved bortkomst ikke erstattes af sygehuset. *practical*

Mere information
Kontakt redaktionen

