

The Feasibility of Using the Semantic Components Model for Indexing Documents in Digital Libraries

*Susan Price

+Marianne Lykke Nielsen

*Lois Delcambre

*Dept. of Computer Science, Portland State University
Portland, OR, USA

+Royal School of Library and Information Science
Aalborg, Denmark

Supporting domain experts using domain-specific digital libraries

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

Our approach

- Leverage knowledge of domain experts using a domain-specific digital library
 - of the types of documents available
 - of the kinds of information in the documents
 - by allowing users to specify search using domain-specific components of documents (not necessarily structural)
- Index documents accordingly

Kolorektalcancer

Disease or Condition

Focus

Epidemiology

6.1.3. Juli 2004

I Roskilde Amt påvises årligt ca 115 tilfælde af kolorektalcancer (KRC), som fordeler sig på 80 koloncancer og 25 rektumcancer - svarende til, at en praktiserende læge i amtet møder knapt ét tilfælde om året.

65% af alle cancerne findes inden for sigmoideoskopets rækkevidde, d.v.s. venstre kolon og rektum. KRC er hyppigst mellem 60-80 år, sjældent før 40-årsalderen. Ca. 5% udgøres af hereditær non-polypøs KRC, hvilket skal mistænkes i familier med flere tilfælde af KRC, heraf mindst én under 50 år.

Symptomer

Disse kan være vage og ukarakteristiske. Typiske symptomer kan være vekslende afføring, blod eller pus i afføringen, vægttab, maven og evt. smerter. Anæmi kan være eneste symptom. Almene symptomer kan være træthed, madlede og vægttab.

Evaluation

Undersøgelser

Alle patienter udredes med almindelig objektiv undersøgelse, abdominal palpation, rektaleksploration, gynækologisk undersøgelse hos kvinder, hæmoglobulin og eventuel ano-/rektoskopi. Familieanamnesen er vigtig. Herefter henvises til sigmoideoskopi.

Herefter henvises til

Referral Guidelines

Servicemål

Amtets servicemål er regeringens udmeldte tidsfrister på 14+14+14 dage.

Visitation af kolorektalcancer

Ved mistanke om KRC henvises patienterne til endoskopisk afsnit på RASK. Ved biopsiverificeret KRC henvises patienterne direkte til kræftsektionen, kirurgisk afd., RAS. (se udredningsprogram - dialogmeddelelse 6.1.2.)

Udredning i sygehusregi samt stadie-inddeling

Efter diagnostisk påvisning af KRC udføres den videre udredning ved kræftsektionen, RAS. Formålet er at stadieinddele canceren og derefter udvælge den korrekte behandling.

Koloncancer

Ultralyd af lever og røntgen af thorax. Herefter stillingtagen til og gennemførelse af operation efter patientinformation og samtykke.

Management Guidelines

Rectumcancer (op til 15 cm fra anokutane overgang)

Rektaleksploration og GU, rektoskopi med biopsi, MR-skanning af det lille bækken, ultralyd af lever og røntgen af thorax. Herefter stillingtagen til og gennemførelse af strålebehandling og/eller operation efter patientinformation og samtykke.

KRC stadieinddeles efter Dukes klassifikation: type A (T1, T2), lokaliseret til slimhinden, type B (T3, T4, nedvækst i eller gennem lamina muscularis (T3) eller til naboelement, knude eller peritoneum (T4)) og type C (alle T-stadier, N1-2), positive

Our approach

- Supplemental indexing that allows search within segments of documents
- Orthogonal to other indexing techniques
 - Full text indexing
 - Keyword indexing
 - Subject description
 - Other metadata

Setting

- sundhed.dk: national Danish health portal
 - Serves needs of clinicians and citizens
 - 24,000+ documents
 - In use since 2001
 - Uses full text and keyword indexing
 - ICPC
 - ICD-10
 - custom thesaurus with lay terms
 - “free” terms
 - Existing vocabularies don’t cover all the information needs and topics of documents in the portal

Outline

- Introduction
- Semantic components
- Overall project
- Indexing study
- Preliminary results

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

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

Document classes and semantic components in sundhed.dk

- Clinical problem – e.g. disease, symptom
 - General information, diagnosis, referral, treatment
- Clinical method – e.g. surgical operation, lab test, radiologic procedure
 - General information, practical information, referral, risks, aftercare, expected results
- Services (patient rights, services provided by healthcare system)
 - General information, practical information, referral
- Clinical unit (hospital specialty department, administrative unit)
 - Function and specialty, practical information, referral, personnel and organization
- Drug
 - General information, practical information, target group, effect, side effects/interactions/contraindications
- Notice or announcement
 - General information, practical information, qualifications

Semantic components

- Some components correspond to facets of the document class
 - e.g. *diagnosis, treatment* of **clinical problem**
 - Content may contain locally-specific information
- Some components group together multiple facets
 - e.g. dose, route of administration in *practical information* component of **drug**
- Some components are more like metadata
 - e.g. *location, responsible official, date* in **environmental analysis** (natural resource management collection)
- Some components contain information specific to collection/user environment, not really facets of topic
 - e.g. *practical information* (where to go), *aftercare* (length of hospitalization, follow-up appts) in **clinical method**

Outline

- Introduction
- Semantic components
- Overall project
- Indexing study
- Preliminary results

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 types and semantic components be identified for a particular domain-specific document collection?

Outline

- Introduction
- Semantic components
- Overall project
- Indexing study
- Preliminary results

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
- Training: introduction
 - to idea of semantic components
 - to 3 document classes and their semantic components
- Tasks: 12 existing sundhed.dk documents
 - Index 6 documents with SC
 - Index 6 documents with keywords
 - Randomly assigned sequence of indexing methods and documents
- Data collection:
 - Indexing data (on paper to avoid UI issues)
 - Time
 - User ease, confidence, satisfaction, and feedback (questionnaires)

Semantic component indexing

ERCP - Endoskopisk Retrograd Cholangio Pancreaticografi

Mere information

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

Kontakt redaktionen

Hvor foregår undersøgelsen?

Undersøgelsen foregår på røntgenafdelingen.

practical

Hvad er en ERCP?

Til ERCP bruges en tynd bøjelig 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).

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

preparation

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 første sal, hvor undersøgelsen vil foregå. Inden undersøgelsen bliver du lokalbedøvet i svelget med en spray og gennem en kanyle i armen får du afslappende og smertestillende medicin. Under undersøgelsen skal du ligge på et røntgenleje i Natostilling (mave/sideleje).

Gastroskopet 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.

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

risik

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.

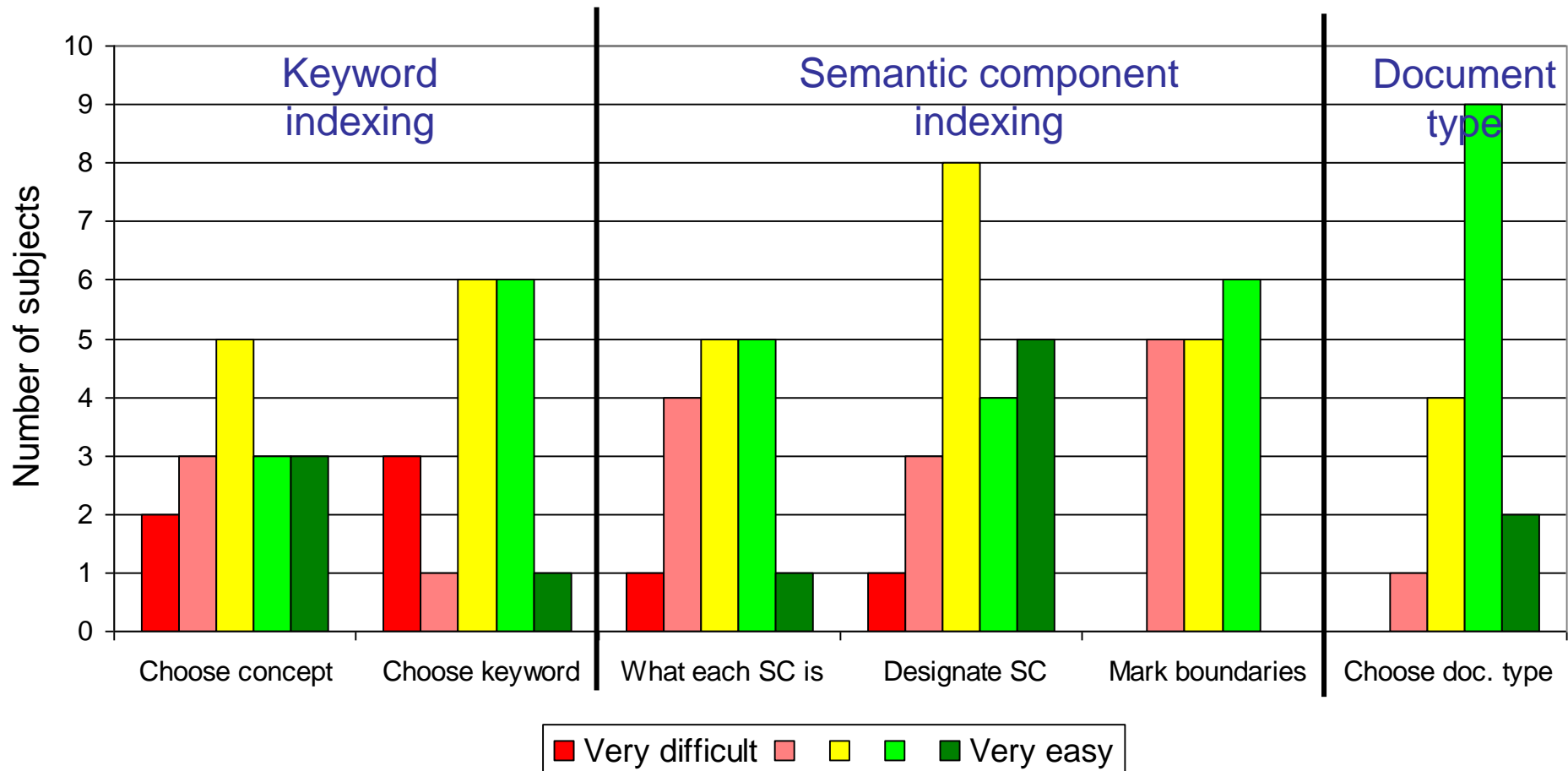
practical

Indexing study: Preliminary results

- Preliminary results:
 - Indexer perceptions and opinions
- Not yet analyzed:
 - Indexing consistency and accuracy
 - Time

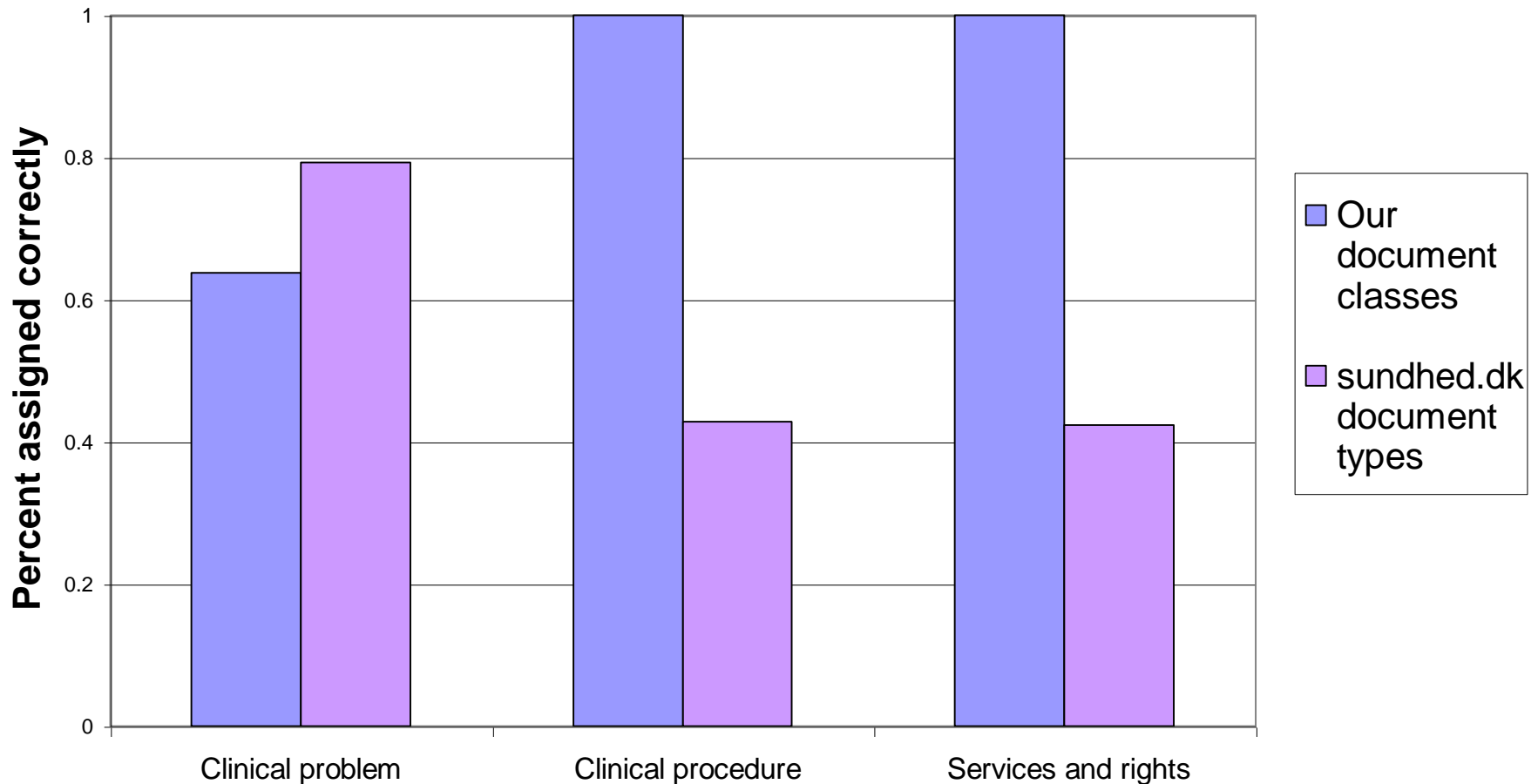
Indexing study results

Indexing difficulty



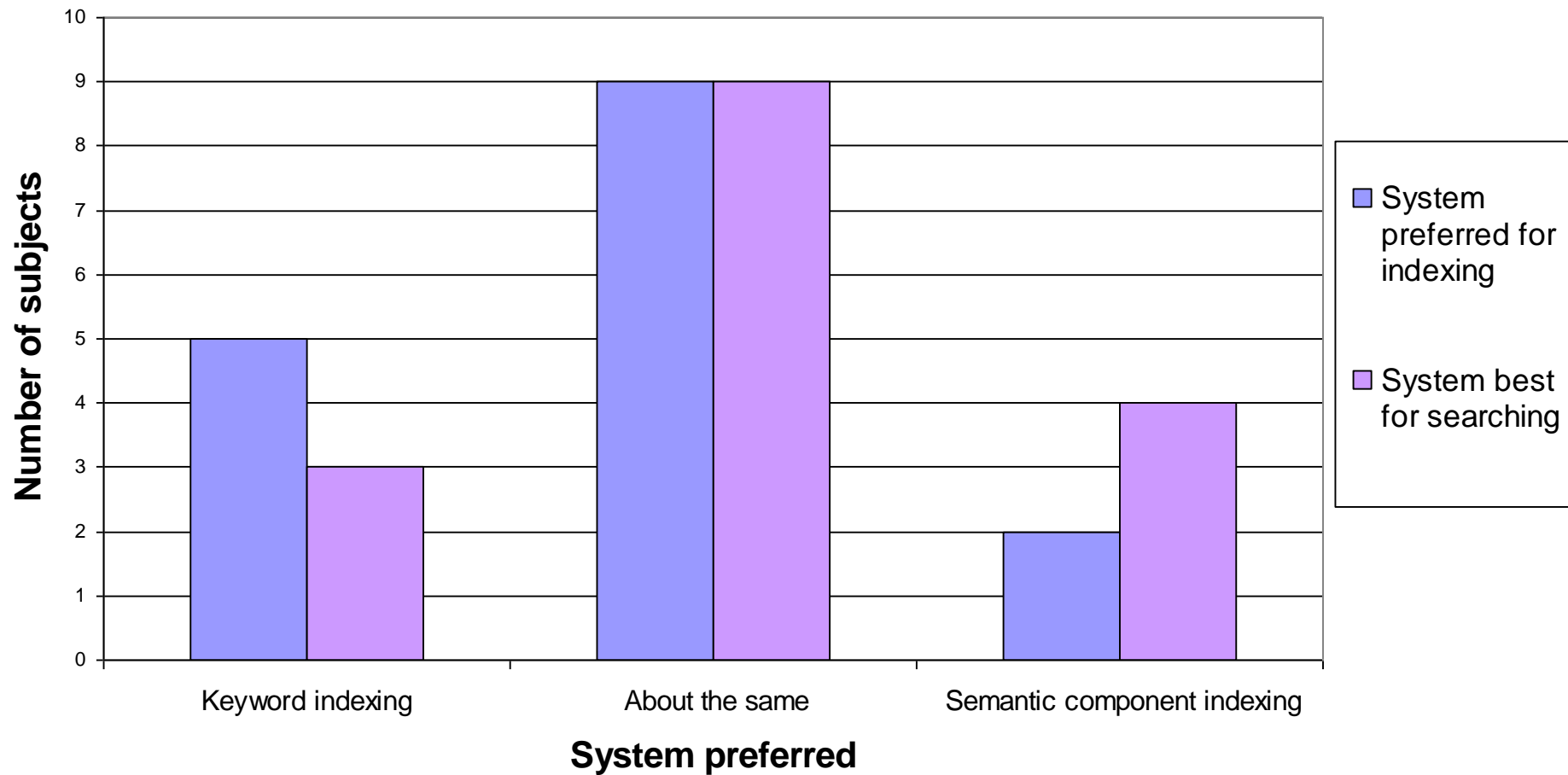
Indexing study results

Document type assignment



Indexing study results

System preferences



Document

(1) Enter Sundhed.dk document id

(2) Press here to get document

[ventefno](#)

Kolorektalcancer

6.1.3. Juli 2004

I Roskilde Amt påvises årligt ca 115 tilfælde af kolorektalcancer (KRC), som fordeler sig på 80 koloncancer og 25 rektumcancer - svarende til, at en praktiserende læge i amtet møder knapt ét tilfælde om året.

>

>65% af alle cancer findes inden for sigmoideoskopets rækkevidde, d.v.s. venstre kolon og rektum. KRC er hyppigst mellem 60-80 år, sjældent før 40-årsalderen. Ca. 5% udgøres af hereditær non-polypos KRC, hvilket skal mistænkes i familier med flere tilfælde af KRC, heraf mindst én under 50 år.

>

>Symptomer

>Disse kan være vage og ukarakteristiske. Typiske symptomer kan være vekslende afføring, blod eller pus i afføringen, luft i maven og evt. smerter. Anæmi kan være eneste symptom. Almene symptomer kan være træthed, madlede og vægttab.

>

>Undersøgelser

>Alle patienter udredes med almindelig objektiv undersøgelse, abdominal palpation, rektaleksploration, gynækologisk undersøgelse hos kvinder, hæmoglobulin og eventuel ano-/rektoskopi. Familieanamnesen er vigtig. Herefter henvises til sigmoideoskopi.

>

>Servicemål

>Amtets servicemål er regeringens udmeldte tidsfrister på 14+14+14 dage.

>

>Visitation af kolorektalcancer

>Ved mistanke om KRC henvises patienterne til endoskopisk afsnit på RASK. Ved biopsiverificeret KRC henvises patienterne direkte til kræftsektionen, kirurgisk afd., RAS. (se udredningsprogram - dialogmeddelelse 6.1.2.)

>

>Udredning i sygehusregi samt stadietinddeling

>Efter diagnostisk påvisning af KRC udføres den videre udredning ved kræftsektionen, RAS. Formålet er at stadietindele canceren og derefter udvælge den korrekte behandling.

>

>Koloncancer

>Ultralyd af lever og røntgen af thorax. Herefter stillingtagen til og gennemførelse af

Semantic Components

Title

(3) Enter your indexer id

(4) Choose the type for the main topic of the document:

- Klinisk enhed
- Klinisk problem
- Klinisk metode
- Services
- Opslag
- Lægemedler

(5) Click here to start indexing

Document

(1) Enter Sundhed.dk document id

(2) Press here to get document

Kolorektalcancer

6.1.3. Juli 2004

I Roskilde Amt påvises årligt ca 115 tilfælde af kolorektalcancer (KRC), som fordeler sig på 80 koloncancer og 25 rektumcancer - svarende til, at en praktiserende læge i amtet møder knapt ét tilfælde om året.

>65% af alle cancer findes inden for sigmoideoskopets rækkevidde, d.v.s. venstre kolon og rektum. KRC er hyppigst mellem 60-80 år, sjældent før 40-årsalderen. Ca. 5% udgøres af hereditær non-polypos KRC, hvilket skal mistænkes i familier med flere tilfælde af KRC, heraf mindst én under 50 år.

>Symptomer

>Disse kan være vage og ukarakteristiske. Typiske symptomer kan være vekslende afføring, blod eller pus i afføringen, luft i maven og evt. smerter. Anæmi kan være eneste symptom. Almene symptomer kan være træthed, madlede og vægttab.

>Undersøgelser

>Alle patienter udredes med almindelig objektiv undersøgelse, abdominal palpation, rektaleksploration, gynækologisk undersøgelse hos kvinder, hæmoglobulin og eventuel anorektoskopi. Familienamnesen er vigtig. Herefter henvises til sigmoideoskopi.

>Servicemål

>Amtets servicemål er regeringens udmeldte tidsfrister på 14+14+14 dage.

>Visitation af kolorektalcancer

>Ved mistanke om KRC henvises patienterne til endoskopisk afsnit på RASK. Ved biopsiverificeret KRC henvises patienterne direkte til kræftsektionen, kirurgisk afd., RAS. (se udredningsprogram - dialogmeddelelse 6.1.2.)

>Udredning i sygehusregi samt stadie-inddeling

>Efter diagnostisk påvisning af KRC udføres den videre udredning ved kræftsektionen, RAS. Formålet er at stadieinddele canceren og derefter udvælge den korrekte behandling.

>Koloncancer

>Ultralyd af lever og røntgen af thorax. Herefter stillingtager

operation efter patientinformation og samtykke.

>Rectumcancer (op til 15 cm fra anokutane overgang)

Generel information

Diagnosticering

Hervisning

Behandling

Semantic Components

(6) Click here when you have finished indexing this document

Generel information

Erase this instance of Generel information

I Roskilde Amt påvises årligt ca 115 tilfælde af kolorektalcancer (KRC), som fordeler sig på 80 koloncancer og 25 rektumcancer - svarende til, at en praktiserende læge i amtet møder knapt ét tilfælde om året.

>65% af alle cancer findes inden for sigmoideoskopets rækkevidde, d.v.s. venstre kolon og rektum. KRC er hyppigst mellem 60-80 år, sjældent før 40-årsalderen. Ca. 5% udgøres af hereditær non-polypos KRC, hvilket skal mistænkes i familier med flere tilfælde af KRC, heraf mindst én under 50 år.

Diagnosticering

Erase this instance of Diagnosticering

>Visitation af kolorektalcancer

>Ved mistanke om KRC henvises patienterne til endoskopisk afsnit på RASK. Ved biopsiverificeret KRC henvises patienterne direkte til kræftsektionen, kirurgisk afd., RAS. (se udredningsprogram - dialogmeddelelse 6.1.2.)

>Udredning i sygehusregi samt stadie-inddeling

>Efter diagnostisk påvisning af KRC udføres den videre udredning ved kræftsektionen, RAS. Formålet er at stadieinddele canceren og derefter udvælge den korrekte behandling.

Hervisning

Erase this instance of Hervisning

Herefter henvises til sigmoideoskopi

Ved mistanke om KRC henvises patienterne til endoskopisk afsnit på RASK. Ved biopsiverificeret KRC henvises patienterne direkte til kræftsektionen, kirurgisk afd., RAS. (se udredningsprogram - dialogmeddelelse 6.1.2.)

Behandling

Erase this instance of Behandling

Additional experience with semantic component indexing

- Indexing to support searching study
 - 371 documents indexed by 6 indexers
 - Used electronic interface
 - Time to index
 - Range 6 sec to 60 min
 - Average = 3 ½ minutes
 - Will analyze further

Future work

- Data analysis of indexing study
- Investigate ways to measure consistency among instances of semantic component indexing
- Investigate methods of automated (or semi-automated) identification of semantic component instances

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.

Acknowledgments

- National Science Foundation
 - International Digital Government Project, Grant #0514238
- National Library of Medicine
 - NLM Training Grant 5-T15-LM07088
- Peter Vedsted MD, PhD University of Århus
- Jens Rubak MD, praxis.dk
- Frans la Cour, Verity
- The sundhed.dk indexers

Thank You