NTNU – Trondheim
Norwegian University of
Science and Technology

Exploring User Interfaces for Search and Content Based Clinical Decision Support in Electronic Health Record Systems

Alexander Perry

Master of Science in Informatics
Submission date: June 2013
Supervisor: Øystein Nytrø, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

Both electronic health records (EHR) and clinical decision support (CDS) are each important attributions to clinicians and clinical workflow. Electronic health records provide clinicians with crucial patient information at the point of care, while clinical decision support gives well-founded and well-documented clinical recommendations at the point of decision making.

This thesis explores how patient information from EHRs could be utilized as a basis for better and more effective decision support. Additionally, two methods of accessing decision support recommendations are created and studied. One is based on search, with known elements from common general and medical search interfaces. The other performs automatic ranking of relevant recommendations based on patient information from a popular norwegian EHR system.

To find out how these two prototypes should integrate and utilize information from an EHR system, a case-based experiment was conducted. This is a qualitative measure of user feedback, with elements from quantitative research. User satisfaction were measured by using methods such as user testing, interviews and surveys.

User feedback suggest that clinical workflow have much to gain from integrating EHR with CDS, as well as computerizing and making clinical national recommendations available in an EHR context. It is also clear that elements like search and automation are important features in an integrated system, and further research of decision support may include an integration of these elements.

Sammendrag

Både elektroniske pasientjournaler (EPJ) og klinisk beslutningsstøtte er hver for seg viktige bidrag til klinikere og klinisk arbeidsflyt. Elektroniske pasientjournaler gir klinikere kritisk pasientinformasjon til rett sted og rett tid, og klinisk beslutningsstøtte gir velbegrunnede og godt dokumenterte kliniske anbefalinger på beslutningspunktet.

Denne masteroppgaven omhandler hvordan pasientinformasjon fra EPJ kan benyttes som grunnlag for bedre og mer effektiv beslutningsstøtte. I tillegg er to metoder for å tilgjengeliggjøre beslutningsstøtte-anbefalinger opprettet og studert. Den ene er basert på søk, med kjente elementer fra felles allmenne og medisinske søkegrensesnitt. Den andre utfører automatisk rangering av relevante anbefalinger basert på pasientinformasjon fra et populært norsk EPJ-system.

For å finne ut hvordan disse to prototypene best kan integreres og utnytte informasjon fra en EPJ-system, ble et case-basert eksperiment utført. Dette er et kvalitativ mål av tilbakemeldinger fra brukerne, med elementer fra kvantitativ forskning. Brukernes tilfredshet ble målt ved hjelp av metoder som brukertesting, intervjuer og spørreundersøkelser.

Tilbakemeldinger fra brukerne tyder på at klinisk arbeidsflyt har mye å tjene på å integrere EPJ med beslutningsstøtte, samt digitalisering av anbefalinger og å gjøre kliniske nasjonale anbefalinger tilgjengelige i kontekst av elektroniske pasientjournaler. Det er også klart at elementer som søk og automatisering er viktige funksjoner i et integrert system, og videre forskning for beslutningsstøtte kan omfatte integrering av disse elementene.

Preface

This master thesis is the result of my master degree in Human Computer Interaction (HCI) at the Norwegian University of Science and Technology. The thesis utilizes HCI in an interesting domain; health informatics. It was written as a part of the EviCare study, with help from DIPS AS.

I would like to thank my supervisor Øystein Nytrø for sharing his knowledge of and experience from health informatics, and Terje Rødsand for providing guidance and expertise regarding alternative usability test methods. DIPS, the Norwegian Knowledge Centre for the Health Services and others in the EviCare project gave valuable feedback and advice, and I submit my gratitude to them as well.

Friends and family also helped with their ongoing support, which is greatly appreciated.

Contents

List of Figures

Acronyms

CDSS Clinical Decision Support System

CDSSs Clinical Decision Support Systems

CDS Clinical Decision Support

EPJ Electronic Patient Journal

DIPS Distribuert Informasjons- og Pasientdatasystem i Sykehus

AGR Automatic Guideline Ranking

GAS Guideline Access using Search

EBM Evidence Based Medicine

EHR Electronic Health Record

NSEP National Centre for Electronic Patient Records

SUS System Usability Scale

RTA Retrospective Think Aloud

IE6 Internet Explorer 6

IE Internet Explorer

HTML Hyper Text Markup Language

CSS Cascading Style Sheets

NOKC Norwegian Knowledge Centre for the Health Services

GCS Google Custom Search

GUI Graphical User Interface

XML Extensible Markup Language

YQL Yahoo Query Language

API Application Programming Interface

SQL Structured Query Language

SIGN Scottish Intercollegiate Guidelines Group

Chapter 1

Introduction

This thesis is part of a larger research project called EviCare (2013), which focuses on integrating evidence-based medicine in clinical information systems. Contributors include Innlandet hospital trust, Norwegian Knowledge Centre for the Health Services, Norwegian University of Science and Technology (Department of Computer and Information Science), Datakvalitet AS and Distribuert Informasjons- og Pasientdatasystem i Sykehus (DIPS). EviCare's main goal is to

> develop methods and technology for providing Evidence Based Medicine (EBM) at the point of care, integrated with an Electronic Health Record (EHR) or other health information systems directly involved in the clinical process, resulting in higher quality of care and a more detailed, transparent documentation of care processes. (EviCare, 2009)

Other projects under the EviCare umbrella focus on parsing and tagging of Electronic Patient Journal (EPJ)s with semantic metadata, while this project explores how this data is best used to support clinicians where and when they make their clinical decisions (clinical decision support).
This thesis will focus more on one of EviCares subgoals, which is to create *practical and usable guideline interfaces - both to healthcare providers and to patients* (EviCare, 2009). More specifically, it will focus on creating varying prototype modules integrated in a clinical information system developed by DIPS. The purpose is to present a set of relevant national clinical recommendations based on selected patient information. This patient information is parsed by other systems in the EviCare project, which then defines recommendation relevance and presents the results to the prototypes in this project.

1.1 Context

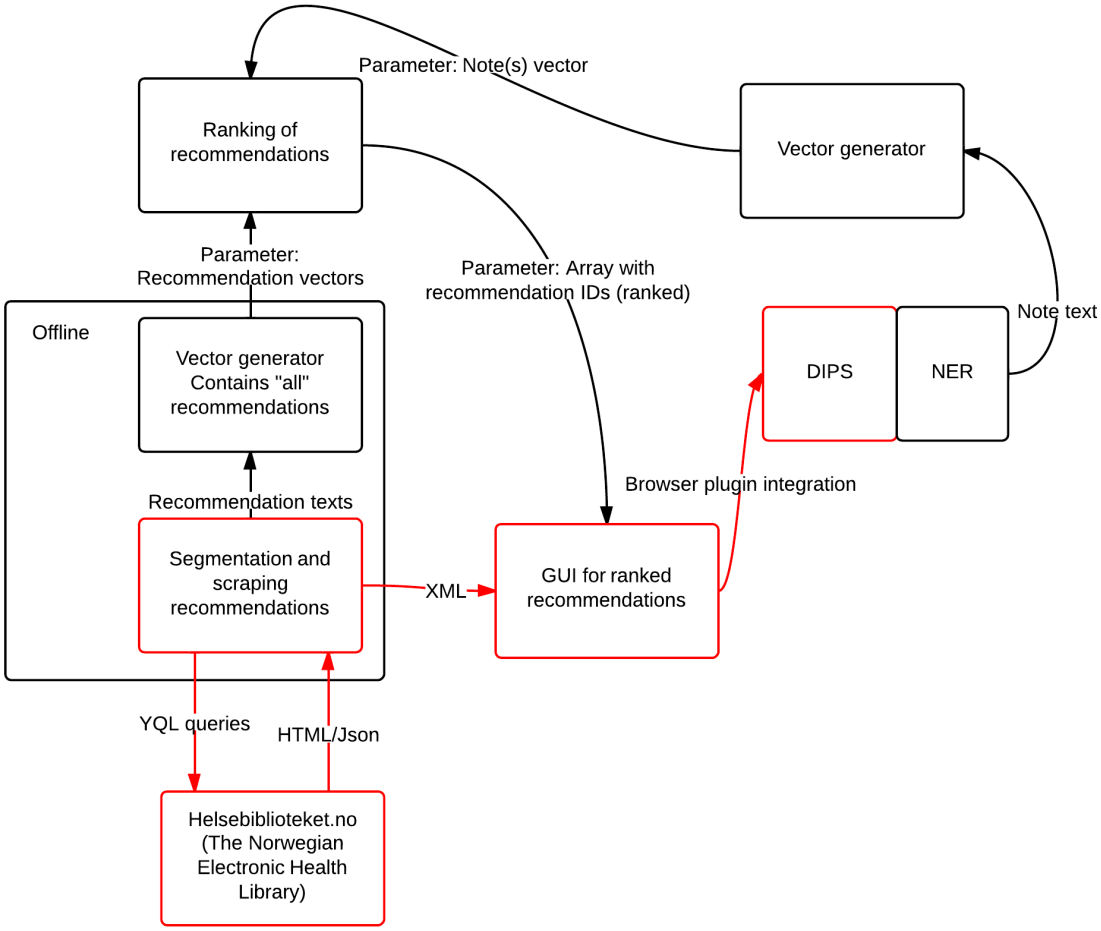


Figure 1.1: Overview over this and adjacent projects

Figure 1.1 represents an overview over the project and adjacent projects under the EviCare umbrella. Boxes and arrows marked in red belongs to this project. An offline script scrapes[1] guidelines from a webside controlled by Norwegian Knowledge Centre for the Health Services (NOKC) called Helsebiblioteket.no (or The Norwegian Electronic Health Library). The scraped results is stored in Extensible Markup Language (XML) format and sent to a prototype website which presents them in ranked order in a Graphical User Interface (GUI). This prototype is then integrated into the DIPS EHR system, and this integrated system is subject to usability testing and evaluation.
Other notable projects include tagging of electronic patient notes (or electronic

[1]Scraping is *the action of taking content from one web site for use un another*

patient records), vector generating from these notes, vector generating based on recommendations, and actual ranking of recommendations. These ranked recommendations are presented in a array with recommendation IDs, which is compatible with prototypes in this project.

1.2 Report Outline

Following is a presentation of the different chapters with a consise description.

Background

Describes the underlying terminology and relevant research projects in the different domains this thesis visits.

Research Method

Works as a baseline for the research conducted in this project, with a description of the chosen research method.

Problem Definition

Defines concrete research questions based on a problem description.

State of the Art

Outlines the current sitation in domains relevant to this master project.

Experiment Design

Specifies research method by defining a concrete research plan.

Prototypes

Provides a description of the prototypes created, and the implementation of these prototypes.

Experiment Execution

This chapter further elaborates on the experiment by describing the execution in detail.

Results

Lists all relevant data results gained from the research.

Discussion

Evaluates and analyse the results in the preceding chapter.

Limitations

Lists and describes key prototype and research method limitations.

Conclusion

Links research questions with discussion, evaluates the project outcome and process.

Future Work

Contains recommendations for future research work, with focus on the prototypes.

Chapter 2

Background

The following background research and definitions provides a base for further exploration of clinical decision support in electronic health record systems. First, electronic health record systems (EHR) are presented and defined, then clinicial decision support systems are explained, followed by an description of clinical guidelines (used for clinical decision support).

2.1 Electronic Health Record Systems (EHR)

Gunter and Terry (2005) defines electronic health record systems as *longitudinal collection of electronic health information about individual patients and populations*. These records often include patients medical history, laboratory results, and personal information like gender, date of birth, social security number and weight. This information is collected from paper and electronic medical records and usually made available for clinicians at hospitals and clinics through networked information systems. These networks range from small and simple networks with few hospitals to wide-range networks on a national basis.

2.2 Clinical Decision Support (CDS)

Clinical decision support (CDS) help clinical workflow by delivering decision support at the point of decision making, to improve patient care. (Garg et al., 2005) Dr. Robert Hayward of the Centre for Health Evidence suggest that clinical decision support may be the missing link between better health information and better health. (Hayward, 2004)

Hunt et al. (1998) defines Clinical Decision Support as

> Electronic or non-electronic system designed to aid directly in clinical

> decision making, in which characteristics of individual patients are used to generate patient-specific assessments or recommendations that are then presented to clinicians for consideration

Earlier studies of Clinical Decision Support Systems (CDSSs) are primarly quantitative studies of the effect of such systems on patient care (Garg et al., 2005), or the actual Clinical Decision Support System (CDSS) itself. (Berlin et al., 2006) They often leave out or neglect qualitative measures like user satisfaction (where the user is either the patient or the clinician) and the clinician's competence level. Kawamoto et al. (2005) did a systematic review of randomised controlled trials of clinical decision support systems, to identify features which improves clinical practice. Findings suggest that the following features significally improves patient care:

- Computerizing clinical decision support
- Delivering decision support at the point of decision making
- Automatic integration of decision support in clinical workflow
- Provide actionable recommendations

It is clear that organizing, collecting and presenting clinical information in an effective and usable manner enables higher quality clinical workflow. It will also most certantly improve information flow, and availability. Another notable finding is that systems with automatic deliverance of clinical decision support had higher success rates than systems where clinicians had to seek out information by themselves. The link between automatic decision support and effective decision support systems is so strong that Kawamoto et al. (2005) suggest implementation of such features in systems wherever it is possible. Todays case is however so that clinical decision support is primarly done through manual systems, or computerized systems with no automated workflow integration and absent emphasis on usability.

2.3 Clinical Guidelines

Clinical guidelines are designed to give clear and concrete recommendations without any form of bias by the source. Guidelines form a link between clinical research and clinical practice by adding evidence-based care to clinical workflows. A committee to advise the public health service in USA have defined clinical guidelines as *systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances.* The primary method of integrating clinical guidelines into patient care is often by analog

methods like through libraries, or through online services like Helsebiblioteket.no. Helsebiblioteket.no is a initiative taken by the NOKC, which develop norwegian clinical recommendations using EBM and makes them available at the point of care. NOKC was established in 2004 and is organized under the Norwegian Directorate of Health. (The Norwegian Knowledge Centre for the Health Services, 2004) Recommendations are based on revised and aggregated research results. Quality measures of this research include systematical review of different relevant sources.

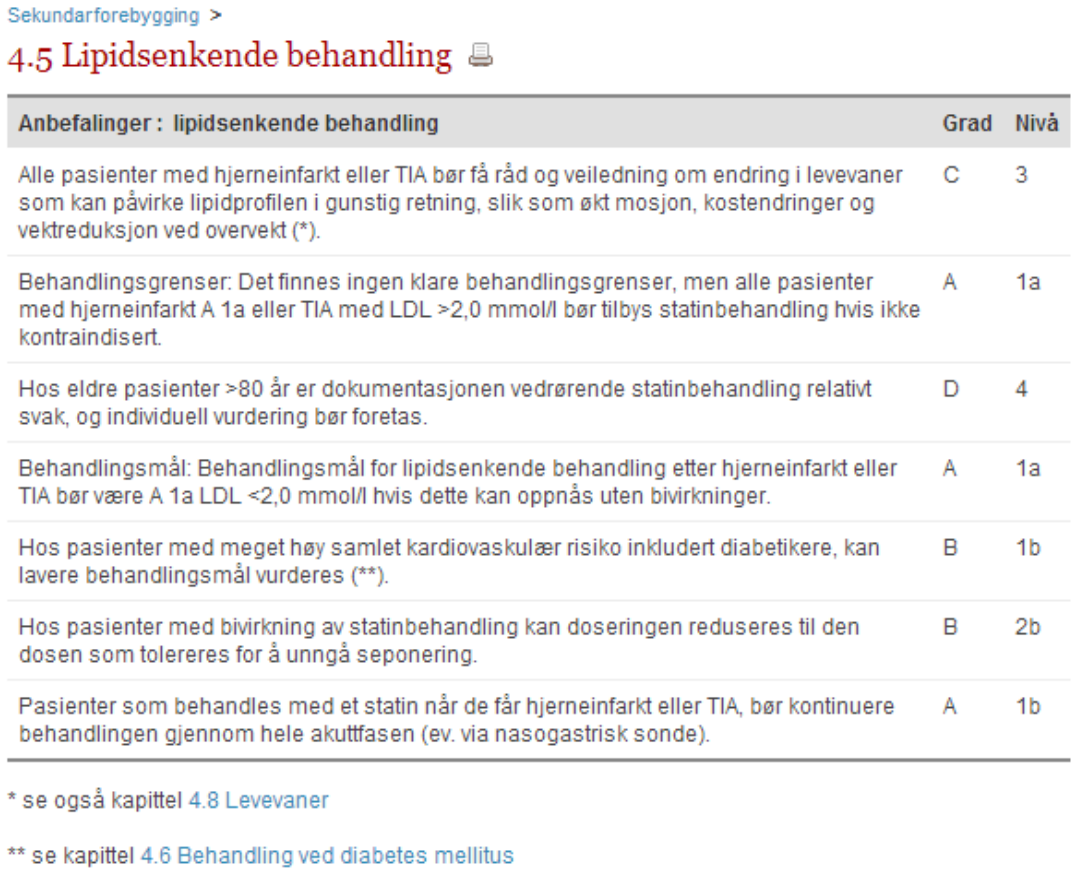
2.3.1 Stroke Guidelines

This project will base the clinical decision support on the Norwegian national stroke recommendations found at Helsebiblioteket.no (link). These guidelines are chosen based on the well-documented research groundwork from Helsebiblioteket. Stroke recommendations and guidelines are based on two national guideline databases and supplementing research results. (Hill, 2008; for Chronic Conditions , UK) The selected group of guidelines is even more specific, only guidelines relevant to secondary treatment of stroke patients are used. These guidelines are less controversial than several of the other national norwegian guidelines.

2.3.2 Grading

The grading of recommendations are based on a grading model developed by a scottish guideline organization called Scottish Intercollegiate Guidelines Group (SIGN) (Healthcare Improvement Scotland, 2013). Knowledge basis is graded per recommendation, on a level basis. Figure 2.3 and 2.2 shows the breakdown of this grading system, originally from Helsebiblioteket (2013) and translated into English.

The different grades are developed by multiple work groups comprised of clinical experts, which have evaluated the documentation strength with ethical, political and economical circumstances in mind.



Sekundarforebygging >

4.5 Lipidsenkende behandling

Anbefalinger : lipidsenkende behandling	Grad	Nivå
Alle pasienter med hjerneinfarkt eller TIA bør få råd og veiledning om endring i levevaner som kan påvirke lipidprofilen i gunstig retning, slik som økt mosjon, kostendringer og vektreduksjon ved overvekt (*).	C	3
Behandlingsgrenser: Det finnes ingen klare behandlingsgrenser, men alle pasienter med hjerneinfarkt A 1a eller TIA med LDL >2,0 mmol/l bør tilbys statinbehandling hvis ikke kontraindisert.	A	1a
Hos eldre pasienter >80 år er dokumentasjonen vedrørende statinbehandling relativt svak, og individuell vurdering bør foretas.	D	4
Behandlingsmål: Behandlingsmål for lipidsenkende behandling etter hjerneinfarkt eller TIA bør være A 1a LDL <2,0 mmol/l hvis dette kan oppnås uten bivirkninger.	A	1a
Hos pasienter med meget høy samlet kardiovaskulær risiko inkludert diabetikere, kan lavere behandlingsmål vurderes (**).	B	1b
Hos pasienter med bivirkning av statinbehandling kan doseringen reduseres til den dosen som tolereres for å unngå seponering.	B	2b
Pasienter som behandles med et statin når de får hjerneinfarkt eller TIA, bør kontinuere behandlingen gjennom hele akuttfasen (ev. via nasogastrisk sonde).	A	1b

* se også kapittel 4.8 Levevaner

** se kapittel 4.6 Behandling ved diabetes mellitus

Figure 2.1: Screenshot of recommendations for cholesterol lowering treatment, from Helsebiblioteket.no

Level	Description
1a	Knowledge based on systematic summaries of randomized controlled trials
1b	Knowledge based on at least two randomized controlled trials, alternatively one large randomized study
2a	Knowledge based on at least one well designed study without randomization, but with an adequate control group
2b	Knowledge based on another type of well designed quasi-experimental study with an adequate control group
3	Knowledge based on well designed, non-experimental, descriptive studies like comparative studies, correlation studies and case studies
4	Knowledge based on clinical experience and consensus in work groups, when there is no sufficient knowledge from relevant studies

Figure 2.2: Grading system for knowledge basis used in guidelines, from Helsebiblioteket.no

Grade	Description
A	Based on excellent documentation with clear results/little risk of bias (level 1a and 1b) and broad consensus in the work group. Grade B is normally used at high risk of bias
B	Based on documentation from at least one good study at level 2a or 2b with low risk of bias or at level 1 with high risk of bias and broad level of agreement/consensus in the work group
C	Based on documentation from well designed non-experimental studies at level 3 or studies at level 2 with high risk of bias and broad consensus in the work group
D	Based on a broad consensus in the work group where no relevant studies of satisfying quality exists

Figure 2.3: Grading of documentation used in guidelines, from Helsebiblioteket.no

2.3.3 Key Features

Guideline use does not guarantee increased quality in patient care. In order to achieve high quality care, recommendations must be consise, non-conflicting, accepted by clinicians/the users, and integrated into the clinical workflow. (Boxwala et al., 2001) Integration into the clinical workflow could be solved by integrating guideline browsing and/or search interfaces into CDSSs. Main reasons for guidelines with insignificant influence on patient care quality include (but are not limited to): (Cabana et al., 1999)

1. Lack of awareness

   clinicians are unaware of the guideline's existence.

2. Lack of agreement

   clinicians simply does not agree with given recommendations.

3. Inertia of previous practice

   clinicians trusts their own experience more than a set of recommendations.

Lack of awareness could be linked to how easy the guidelines are to obtain. If clinicians have to discover guidelines almost randomly by browsing through a set of trustworthy sources, then the required effort is possibly to demanding, and clinicians will discard the search.

Lack of agreement are an important point, as not all guidelines are deemed correct by all clinicians. Many (including stroke) guidelines are however based on credible and tested background research with detailed documentation. Guidelines may also produce inconsistency in Clinical Decision Support (CDS) and some guidelines even give conflicting recommendations. Systems will need to be transparent with both background research and documentation, without overflowing clinicians with information.

Inertia of previous practice are somewhat related to the previous point, where sources are deemed non-credible by clinicians. In these situations, clinicians will most likely discard recommendations, and use their own experience.

Chapter 3

Research Method

In research, there is two common methods for data collection and analysis. The first focuses on data based on numbers, or **quantitative** data. Data collection in this research method are primarly done through means of experiments and surveys. (Oates, 2006, chapter 17) Experiments are *strategies which investigates cause and effect relationships, seeking to disprove a casual link between a factor and an observed outcome.* (Oates, 2006, chapter 9) The goal of a survey is to collect a large amount of data in a standarized and systematic way. Generalizable data is often the focus point.
Qualitatative data collection and analysis focus on all non-numeric data such as data from interviews, websites and developer models. (Oates, 2006, chapter 18) Common methods of data obtainment include case studies, and ethnography. Ethnography describes people or cultures and their norms, while case studies focuses on one instance; either an organization, a development project or an information system. Data analysis in qualitative methods may include quantitative analysis methods.
A combination of both qualitative and quantitative methods are used in this research project, by conducting a exploratory case-based experiment. Case-based experiment focuses on a given case, which in this project is a CDSS. But it also features some experiment based methods, with hypothesis and controlled variables. The hypothesis here is that user satisfaction will increase with clinical decision support integrated into electronic health record systems. In addition, theory suggest that clinicians would adopt computerized clinical decision support into their workflows when certain factors are met. Factors include **availability**, **credibility**, and **relevance**.
The experiment is meant to investigate changes in user satisfaction, and perceived availability, credibility and relevance when using different means of acquiring clinical guidelines. An EHR system is the system used in testing, while the variable is an integrated module for guideline obtainment. This module is the independent

variable in the experiment, meaning that it is changed under controlled circumstances, in order to produce different results. Variation lies in navigation, where one variant is to be implemented with a search-based interface, while the other is to be implemented using automatic content-based interaction. Content-based, meaning based on electronic health records from the EHR system. These results are then analysed, and the analysis provides evidence which may confirm or contradict the hypotheses. The purpose of this experiment is to do exploratory work, and then provide recommendations for future tests and future systems based on the resulting data. System Usability Scales are used for supplementing data collection.
Results are defined by self-reporting data, and data is collected by using usability tests and interviews. The group of participants is as homogenous as possible under these circumstances. They are medical students from the same medical study at NTNU, but with differing practical clinical experience. Recruitment of participants is somewhat at random, with the exception of specifying class year. Their responses constitutes the independent variables, which means variables that can not be controlled by adjusting the circumstances. Experiment design is explained more thoroughly in chapter 6.

Chapter 4

Problem Definition

This chapter further elaborates on findings described in chapter 2, and defines the problem this project sets out to solve in the form of research questions.

4.1 Common Problems

Chapter 2 stated several problem related to clinical decision support, and concluded that clinical decision support is as good as useless without certain features. These features include (but are not limited to) **computerization**, **workflow integration**, and **availability**, and provide a basis for further research in this project.

4.1.1 Workflow Integration

This exploratory research tries to find out how clinical decision support systems best can be integrated in a clinical workflow, and use elements from this workflow to provide decision support.

4.1.2 Computerization

Clinical decision support is widely used, but many solutions depend on manual data collection with one local source of information. Digital solutions does exist, but information here may contradict local practice at the site of decision making. Other computerized information systems does exist, and could provide a basis for a clinical decision support system.

4.1.3 Availability

Navigation is often cumbersome when information sources for recommendations are spread out over analog sources such as libraries, and digital sources such as

differing online databases like UpToDate (2013) and Helsebiblioteket.no.
Different sources for guidelines and recommendations also often use several different means of navigation, which gives clinicians inconsistent interfaces in the tools they rely on. Guidelines and recommendations become unavailable to some extent for clinicians as a consequence of incosistent (and sometimes bad) interaction design. Helsebiblioteket.no is the main source of national guidelines in Norway, and they operate with several different interfaces based on what type of guidelines and recommendations clinicians are seeking.

4.2 Project Description

The goal is to explore how search-based and content-based interfaces to national stroke guidelines can best utilize information such as EHR in an integrated environment. The stroke guideline contains many, fine-grained, recommendations for different problems, different phases of the disease, different patient states and preferences.
The method is to test different levels and modes of user interaction, as well as different rankings and presentations of the recommendations in the stroke guideline. In order to do realistic testing, a web-based interface will be fully integrated in an existing and working EHR system (DIPS) with a comprehensive and truthful patient case. Testing will involve medical students with hospital experience, an will be performed in the IDI (Department of Computer and Information Science) usability lab for health information systems. The hypothesis is that a higher level of user interaction in CDSSs affects user satisfaction in a positive way and, as a consequence, the overall usability will improve.

4.3 Research Questions

The idea is to test integrated CDSS (in an EHR) and see how different variations in user interaction affect users, usability and workflow. These are the main two research questions to be answered in this thesis, with specifying subquestions;

RQ1 How can a CDSS best utilize patient information as a context for ranking relevant recommendations?

- **RQ1.1** Which of search-based or content-based recommendation ranking gives best user satisfaction?

RQ2 How does CDSS integrated in the EHR system affect clinical workflow?

RQ2.1 How does clinical users respond to integrated decision support in the EHR system?

RQ2.2 How does automated computerized decision support compare to manual decision support?

Chapter 5

State of the Art

Chapter 2 introduced terms such as clinical decision support, electronic health records and clinical guidelines. This chapter describes the current *State of the Art*[1] of medical information retrieval methods which utilize such systems and features, with focus on functionality and design.

5.1 Medical Information Retrieval

Medical search engines are widely used, but there are few qualitative studies on the effects of integrating said search engines into CDSSs and guideline retrieval. Helsebiblioteket.no is a collection of Norwegian national guidelines, and provides additional links to other guideline databases. The website utilizes different navigational interfaces such as search and manual browsing. Guidelines for stroke treatment, prevention and rehabilitation are grouped in categories based on time of care. These categories function as chapters, where different guidelines function as subchapters. Figure 5.1 shows the interface for the national stroke guidelines, and the category for secondary preventative guidelines is selected.

[1] Definition of SoA: *the highest level of development at a particular time (especially the present time*

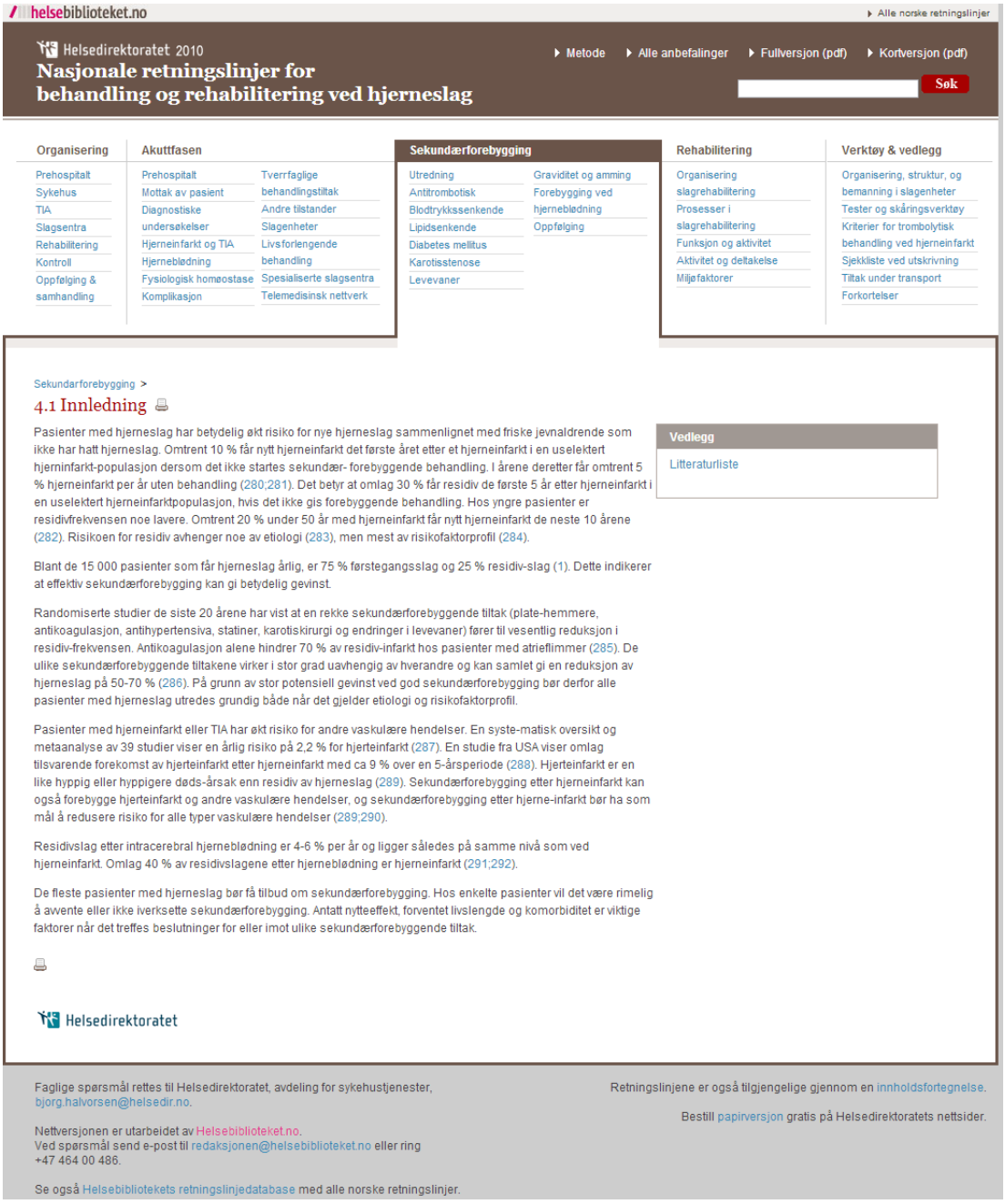


Figure 5.1: Helsebiblioteket.no, stroke guidelines, secondary prevention

5.1.1 Clustering

Clustering is another way of organizing items, where the actual presentation of information is somewhat similar to structured guided browsing and searching.

Items are grouped together in different categories, by how alike the different items are. Clustering are often automatic, and is achieved by analyzing phrases or words. (Cutting et al., 1992) They are then presented in categories relevant to their content. This automation is a clear advantage over structured guided browsing, but clustured information retrieval often involve unclear or illogical categories and labeling. (Baeza-Yates and Ribeiro-Neto, 2011, page 39)

5.1.2 Structured Guided Browsing

Structured guided browsing is also often used in websites like Helsebiblioteket, where different documents are collected in parent categories. Helsebiblioteket does noe utilize structured guided browsing, meaning that the organization of guidelines deviate from other structured hierarchies. Structured guided browsing means browsing through a hierarchy of information using data structures as a guiding tool. Data in such models are often structured in hierarchies, or in tree structures. Directories like Yahoo (2013a) uses structured guided browsing with a tree structure, where parent domains such as *Arts and Humanities* contain child nodes with other sub-domains such as *Photography*, *History*, and *Literature*. A directory is basically a collection of documents, where the documents are collected in parent categories.

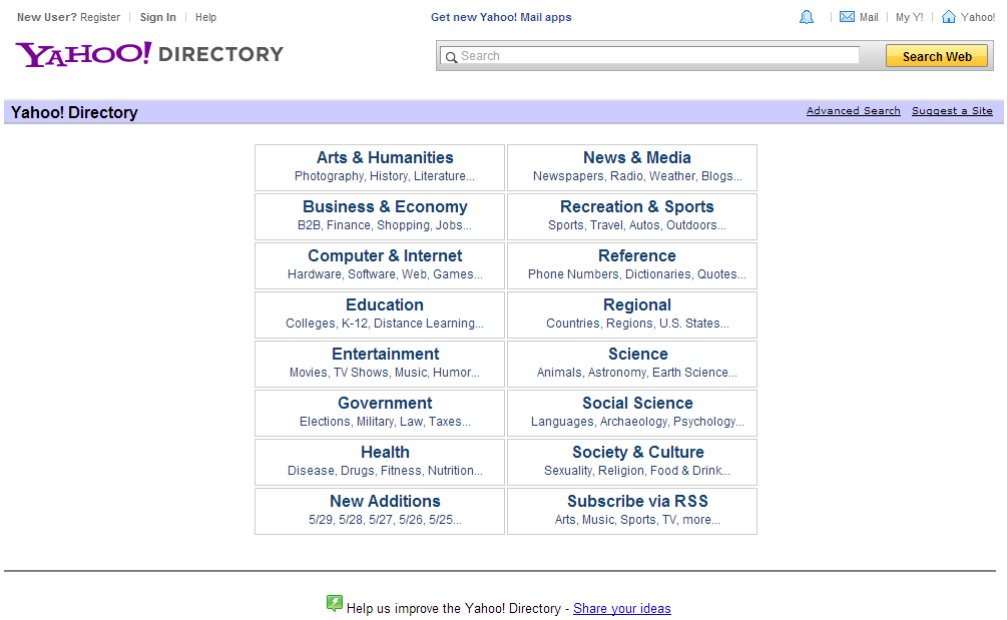


Figure 5.2: Screenshot of Yahoo Directory showing various categories and subcategories

5.1.3 Search

Directories offers precision when searching for information, but offer little to no recall. Low recall is due to low web coverage, and less than 1% of all web pages are covered. (Baeza-Yates and Ribeiro-Neto, 2011, page 499-500). Search based browsing covers a lot more, and directories are often combined with search solutions to improve recall. Search engines are today clear leaders when it comes to information retrieval on the world wide web. Google claimed to have indexed 26 million sites in 1998, which increased to one billion sites in 2000. In 2008, Google hit the one trillion mark, and that is only a part of the whole web. (Alpert and Hajaj, 2008)
A recent survey of nine similar online tools for medical information retrieval found that nephrologists[2] use well known medical information systems like UpToDate (guideline collection) (UpToDate, 2013) and PubMed (search interface to a medical research database) (US National Library of Medicine, 2013) well as general search engines such as Google Search. (Shariff and Garg, 2011) These results show that it is possible to create medical information retrieval solutions by utilizing search as a way of finding relevant information. It is important to note though, that this particular study does not focus on clinician satisfaction and how much the clinicians trust these sources. Medical information retrieval sites could clearly benefit from other large information sources. Google is the clear leader on this front, and the simple yet effective design of Google Search is a significant reason for the site's popularity. The site uses the familiar search rectangle, which has been used almost since day one, and is being used by several other search providers. This design element is dubbed **The Search Rectangle Paradigm** in (Baeza-Yates and Ribeiro-Neto, 2011, page 481-482). With little excessive information, any person is able to find what they want with little to no training. Previously mentioned UpToDate has incorporated this simple design to some degree, while also incorporating topic search, which enables clinicians to define topics for their search queries.

[2]Definition of nephrology: *a branch of medicine concerned with the kidneys*

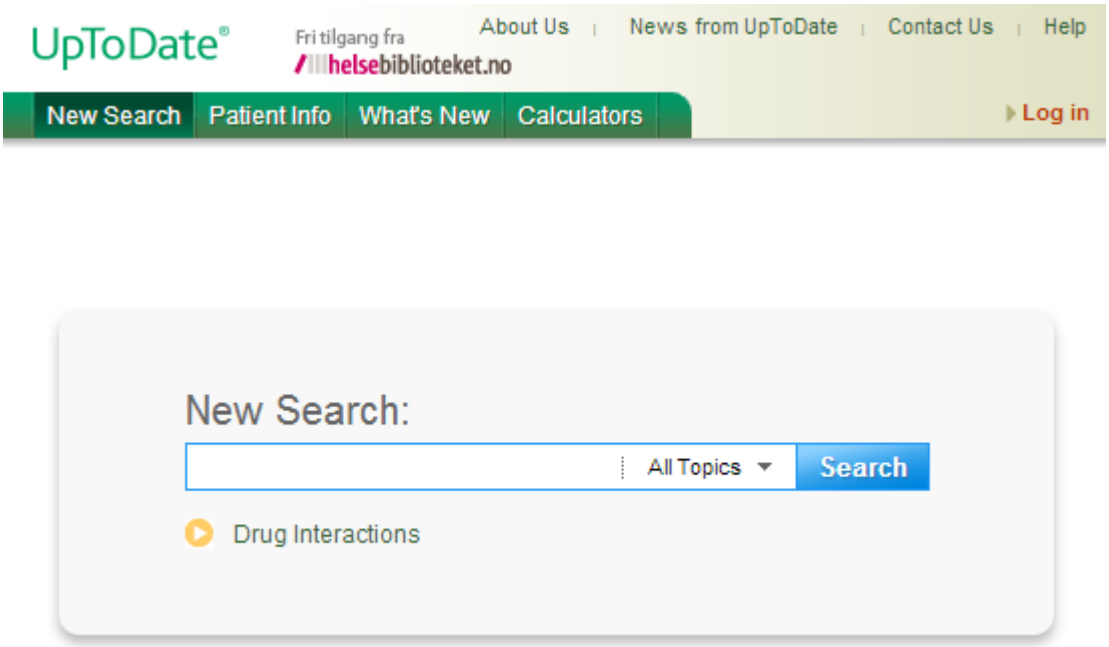


Figure 5.3: UpToDate search interface with a simplistic design

Simple design creates a lower barrier, which is useful as studies show that lack of training is one factor that hinders clinicians when searching for information. (Doney et al., 2005) In order to make a successful medical information source with a search interface, designers of those sites needs to focus on the users. Clinicians are the target audience in this scenario, and they need effective and trustworthy sources. A time consuming information retrieval process with results overflowing with conflicting and thus unreliable information are the main reasons for clinicians skepticism. (Scott and Fairweather, 2000) The clinicians needs relevant information and clinical decision support at the point of care, they normally have little to no time for browsing medical sources at libraries to gain knowledge. (Lappa, 2005)

5.1.4 Context in Browsing

A downside to manual browsing is that users risk losing context when diving deep into different substructures. Helsebiblioteket.no features a element in each chapter called breadcrumbs, which show users the chapter context, with preceding chapters. Breadcrumbs depicts the path from the root node to the selected node, as shown in these screenshots from Windows Explorer in Windows 7 and from Helsebiblioteket.no:

Figure 5.4: Breadcrumbs in Windows Explorer

Figure 5.5: Breadcrumbs at Helsebiblioteket.no

Chapter 6

Experiment Design

The experiment design is based on earlier master theses in the EviCare project by Terje Røsand (2012) and Zheng Wang (2012). Terjes thesis focused on methodology around usability testing with think aloud and eye tracking while Zheng Wangs thesis focused on the actual integration of clinical guidelines and electronic patient journals, and more specifically on requirements in that domain.

6.1 Goal

This experiments goal is to measure user satisfaction of a CDSS integrated with an EHR, and at the same time look at how this integration may affect clinical workflow. Two different means of guideline navigation in clinical decision support are measured against each other mainly through user testing and retrospective interviews, with system usablity scales (SUS) as supplement.

6.2 Electronic Health Record System (EHR)

The chosen EHR system is created and maintained by DIPS, and used in a range of hospitals and clinics in the norwegian health sector, with notable hospitals like St. Olavs Hospital in Trondheim, and Akershus University Hospital. In addition to delivering electronic health records at the point of care, the system also offers a patient administration interface. DIPS EHR enables easier clinical collaboration and patient information access and offers features such as multimedial information and speech recognition for digital dictation.

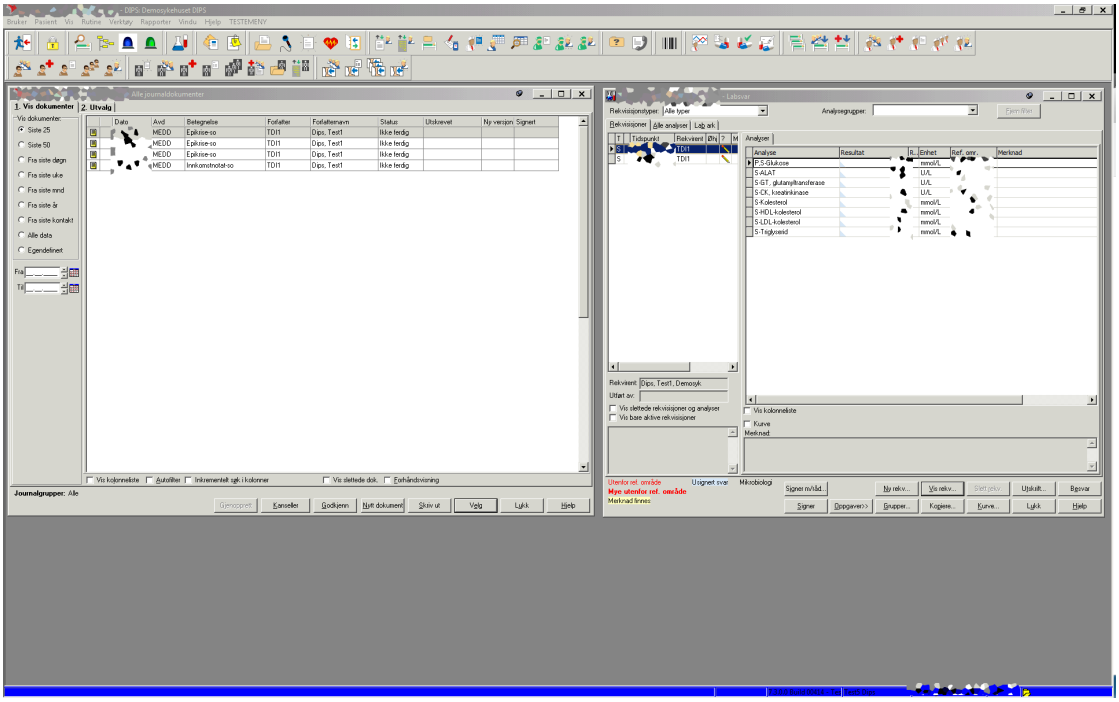


Figure 6.1: Screenshot of DIPS with patient records

6.3 Clinical Decision Support System (CDSS)

The case in this case-based experiment is not the DIPS EHR system, but rather a CDSS module integrated in this system. Two different prototype modules are to be created and integrated into the EHR system. One with search-based navigation, and another content-based system. Search-based means greater user interaction, where clinicians would define their own searches based on health record information in the DIPS EHR system. Content-based means little to no user interaction. Guidelines are presented using data from a ranking algorithm, and presented in ranked order based on the content in the selected patients electronic health record. Both prototypes presents stroke guidelines described in chapter 2.3 to the clinicians. The technical implementation of both guidelines are also described further in chapter 7.

6.4 Target Audience

The DIPS system is intended to be used by hospital clinicians, nurses, physicians and other health personnel required by law to maintain a record of patient care and treatment. As seen in chapter 2, experienced clinicians often weigh their own opinion more than external recommendations like the national guidelines this experiment utilizes. Nevertheless, the target audience for this experiment is recently graduated clinicians, since they are easier to recruit. If experienced clinicians trust their own opinions and experience more, then inexperienced students may have more of an open mind towards new and external input.
Clinicians in also often have tight schedules, and therefore are less likely to partake in such experiments as this. While students often have the same problems with tight schedules, they are more available. They also frequent the hospital buildings where the usability lab is located. Since it is important to recruit clinicians with actual experience from practical work, the target group were adjusted to include only medical students from third grade to sixth grade. These students have also gained sufficient expertise on clinical guideline use from several courses at NTNU.

6.5 Patient Case

To make this experiment as realistic as possible, a patient case with discharge notes, lab values and other relevant information is used. This patient case is refreshed and slightly modified from a previous experiment in the EviCare project. Although only one patient case is used, there is two patient entries in the DIPS system. One patient were called "Stein Henriksen" and one called "Vidar Havnut". The former corresponds with the search-based interface, while the latter corresponds with the content-based list of guidelines. Participant were presented the following patient case before the usability test, in norwegian:

> *Pasienten er en mann født 20.06.1961, og fikk et hjerneslag for to år siden. På Rikshospitalet ble han behandlet for åpen foramen ovale. Han har vært på poliklinikken to ganger tidligere, og dette var det tredje besøket. Hans LDL kolesterolnivå var på 2,4.*
> *I følge retningslinjene, skal alle pasienter som har fått hjerneslag og har et LDL-nivå over 2,0 bli tilbudt statinbehandling, noe som er et lipidsenkende legemiddel.*

6.6 Usability Testing

Usability is defined by Nielsen (2012) as a *quality attribute that assesses how easy user interfaces are to use*. Usability testing is a method of studying usability of a given user interface. The most effective and advantegeous method of usability testing is user testing. User testing involves testing the user interface with users representative for the target audience of the test object. These user tests are often conducted in usability labs, and data collection is done through observation. The usability evaluation in this experiment is done mainly user testing and interviews, and are performed at a usability lab.
A usability test is conducted by giving the user a set of actions/tasks to perform and then observe the user without (or with little) interruption. These tasks are predefined and based on previous observations of clinicians workflow from earlier research projects. The target audience are clinicians, and two different prototype alternatives are created based on the research questions.

6.6.1 Usability

Usability is defined by the International Organization for Standardization in ISO 9241-11 as the *extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use* (Iso, 1998) (Nielsen, 2012)

6.6.1.1 Effectiveness

It is common to measure the effectiveness of a system by measuring the users ability to carry out a given set of tasks. Accuracy is also a common measurement unit for determining the effectiveness.

6.6.1.2 Efficiency

Efficiency is often measured with task completion time; the amount of time the user spend per task. Number of completed tasks in a given time period is also a common method of measurement.

6.6.1.3 User Satisfaction

User satisfaction is an important factor in usability research, and this qualitative way of measuring usability is more relevant to the overall goal in this experiment.

6.6.2 Think Aloud and Retrospective Think Aloud

Think aloud methods are often used in connection with usability testing. Test users are encouraged to verbalize their thoughts while performing tasks. This method were first introduced by Lewis (1982), and later reviewed by Lewis and Rieman. (Lewis and Rieman, 1993). This method is so common due to its effectiveness when it comes data gathering about factors that affects satisfaction, effectiveness and efficiency. Think aloud have to be used in combination with video and audio recording, or note taking (less effective), in order to fully utilize the method.

6.7 Preliminary Survey

Participants were asked to answer a simple preliminary survey, which gathered data about their previous experience in relation to stroke patients and guideline use. It also mapped their age, gender, name and class year. The purpose of this survey is to map prior experience with clinical guidelines and user habits when using clinical decision support. The goal is to see what methods of accessing guidelines participants in the experiment prefer, and how much they use clinical guidelines as decision support in general.

6.8 Variables

Experiments are conducted by altering *independent variables* to see how the *dependent variables* are affected. Independent variables include variables that are under control, such as the CDS prototype modules in this experiment. One module is search-based, and another is content-based (based on ranked guidelines which in turn is based on content) as mentioned in chapter 3. The dependent variables are not controllable, and clinicians self-reporting data is the dependent variables in this experiment. This means that feedback from the interviews, usability tests, surveys and forms act as dependent variables.

6.9 Hypothesis

The hypothesis here is that features with automated decision support such as the content-based solution (with pre-ranked guidelines based on EHR content) gives a better user satisfaction than the search-based one. This is only true if the content-based system inhibits a non-intrusive design and offers what clinicians see as relevant guidelines.

Another hypothesis is that computerized integration between EHR and CDS creates a workflow which clinicians prefer over manual decision support. Manual meaning decision support by means of external digital guideline collections, or accessing physical guideline collections.

Chapter 7

Prototypes

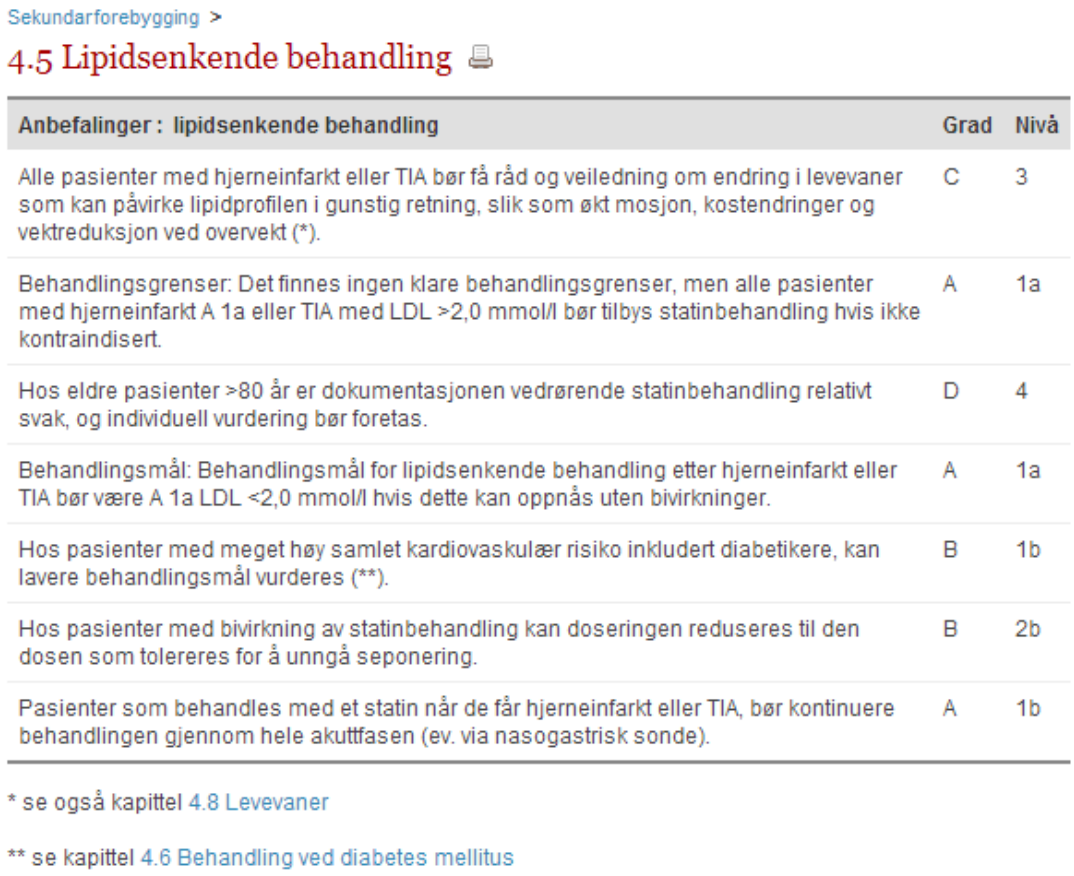
In order to see what differences in guideline navigation does to user satisfaction, it is necessary to test solutions with varying navigational methods. Chapter 5 mentions search and tree structured browsing, two different possible solutions for accessing clinical guidelines. To measure what affect these features have on clinician workflow and user satisfaction, they need to be separated. Therefore, one prototype[1] is created for each feature; one search-based and one content-based.
Another goal of this exploratory research is to see how much users trust automated solutions in integrated clinical decision support. The prototype which provides decision support through browsing receive ranked lists of guidelines. This list is generated based on patient content in EPJs. The prototype use this list to present relevant guidelines.
These two prototypes are to be integrated into a system for EHRs, created by DIPS. Each prototype module is accessed through a browser plugin, and the search module is called **Guideline Access using Search** or **GAS** for short. The module with automated decision support is called **Automated Guideline Ranking** or **AGR** for short. Refer to the list of Acronyms for an overview over this and other acronyms used in this thesis (or just click the acronyms in the text when reading the pdf). On the back-end, a system parses guidelines from Helsebiblioteket.no. (The Norwegian Knowledge Centre for the Health Services (NOKC), 2013)

7.1 Guidelines and Recommendations

Guidelines are parsed from Helsebiblioteket.no as a table, where each row represents a recommendation. The table content is the only consistent element being used for this experiment, presentation and styling is altered. Figure 7.1 shows

[1]Definition of prototype: *an original model on which something is patterned*

the original guideline for cholesterol lowering treatment. Each row in the table represents a recommendation, with grading and level information.



Sekundarforebygging >

4.5 Lipidsenkende behandling

Anbefalinger : lipidsenkende behandling	Grad	Nivå
Alle pasienter med hjerneinfarkt eller TIA bør få råd og veiledning om endring i levevaner som kan påvirke lipidprofilen i gunstig retning, slik som økt mosjon, kostendringer og vektreduksjon ved overvekt (*).	C	3
Behandlingsgrenser: Det finnes ingen klare behandlingsgrenser, men alle pasienter med hjerneinfarkt A 1a eller TIA med LDL >2,0 mmol/l bør tilbys statinbehandling hvis ikke kontraindisert.	A	1a
Hos eldre pasienter >80 år er dokumentasjonen vedrørende statinbehandling relativt svak, og individuell vurdering bør foretas.	D	4
Behandlingsmål: Behandlingsmål for lipidsenkende behandling etter hjerneinfarkt eller TIA bør være A 1a LDL <2,0 mmol/l hvis dette kan oppnås uten bivirkninger.	A	1a
Hos pasienter med meget høy samlet kardiovaskulær risiko inkludert diabetikere, kan lavere behandlingsmål vurderes (**).	B	1b
Hos pasienter med bivirkning av statinbehandling kan doseringen reduseres til den dosen som tolereres for å unngå seponering.	B	2b
Pasienter som behandles med et statin når de får hjerneinfarkt eller TIA, bør kontinuere behandlingen gjennom hele akuttfasen (ev. via nasogastrisk sonde).	A	1b

* se også kapittel 4.8 Levevaner

** se kapittel 4.6 Behandling ved diabetes mellitus

Figure 7.1: Screenshot of recommendations for cholesterol lowering treatment, from Helsebiblioteket.no

7.2 Guideline Access using Search

GAS features a simplistic user interface, inspired by the rectangle paradigm used in Google Search and UpToDate. The rectangle paradigm is mentioned in chapter 5.1. Google Custom Search is used as a basis for the search functionality, due to more effective implementation.

The most effective and feature rich way to implement this is to use a custom search provided by Google.

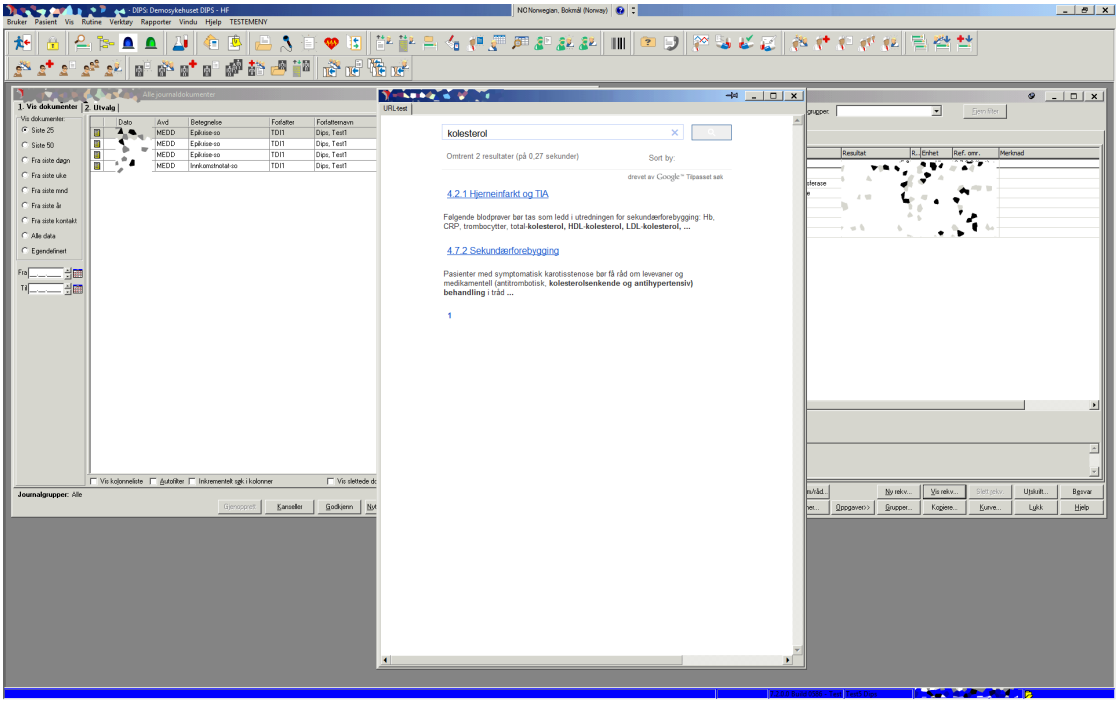


Figure 7.2: Screenshot of DIPS with GAS

7.2.1 Google Custom Search

Google Custom Search (GCS) features a simple search interface, very similar to the official Google Search user interface. It uses the search rectangle paradigm, where the most dominant element in the GUI is an input box for search queries. GCS enables customization through a simple web interface, where developers can do simple stylistic and functionality adjustments. Google Custom Search (2013) The functionality features include (but are not limited to) defining and selecting synonymous terms, selection of sites to crawl[2], and custom autocompletion. In GAS, custom synonyms are used with clinical terms.

[2]Definition of web crawling: *Traversing web sites systematically for the purpose of web indexing*

Figure 7.3: Screenshot of Google Custom Search in GAS

URLs to the guidelines in the result list are blanked out, to remove irrelevant and confusing elements. Advertisements are also obviously disabled, as this would greatly reduce credibility and defy the simplistic design.

7.3 Automatic Guideline Ordering

AGR are a more static module than GAS, only displaying the parsed guidelines in a specific order. The purpose of AGR is to explore how users interact with automated decision support, ie. automatic access to relevant recommendations. Another module in the EviCare project features an algorithm where guidelines are ranked based on tagged EPJs in the DIPS system. AGR requires input in form of an array of recommendation IDs, and the order in the array denotes the guideline order in the module interface.
Adjacent projects ranking algorithms and EPJ tagging are finished in parallell with this project. Therefore, a pre-determined static array list of ranked recommendations is created for testing purposes. Relevance to the patient case are pre-defined by clinicians, and guidelines are ordered as following:

1. Cholesterol lowering treatment (Lipidsenkende behandling)
2. Blood pressure lowering treatment (Blodtrykssenkende behandling)
3. Secondary prevetion of cerebral hemorrhage (Sekundærforebygging ved hjerneblødning)
4. Follow up of secondary prevetion (Oppfølging av sekundærforebygging)

5. Elucidation for secondary prevetion (Utredning for sekundærforebygging)
6. Antithrombotic treatment (Antitrombotisk behandling)
7. Diabetes mellitus and treatment (Behandling ved diabetes mellitus)
8. Carotisstenose and treatment (Behandling ved karotisstenose)
9. Living habits (Levevaner)
10. Secondary treatment, pregnancy and breastfeeding (Sekundærforebygging ved graviditet og amming)

A problem with users browsing through results is that users often lose perception of context, which is discussed further in chapter 5.1. AGR is designed to show contextual information both through chapter headings and breadcrumbs, to combat this issue.

7.4 Integrating modules and DIPS

Decision support are not at all integrated in to the DIPS EPJ system today, and clinicians are forced to either rely their own (or colleagues) experience and knowledge, use static sources like the hospital library, or look up guidelines through several online sources. This affects clinicians workflow to a certain degree, by taking up time which could have been used on patient care, and by reducing motivation to update clinical practice by using modern and verified recommendations. Systems found at sites like PubMed (US National Library of Medicine, 2013) and Helsebiblioteket.no (The Norwegian Knowledge Centre for the Health Services (NOKC), 2013) could easily be integrated into CDSSs like the DIPS system, and therefore be integrated into clinicial workflow.
The DIPS system features a module-based web browser plugin, which is set to open a given web address controlled by National Centre for Electronic Patient Records (NSEP). A PHP-script is located at this web address, and the purpose of this script is to redirect users to either two of the prototype modules. It decides based on a patientID, which is given as a parameter. The two modules are linked to two test patients, where patient one (Stein Henriksen) has a patientID of `1000270` and patient two (Vidar Havnut) has a patientID of `1000218`. The two modules are temporarily hosted at NTNUs servers, for easy access during tests. Test participants are not affected by this, as the browser plugin only displays the actual site, and removes other non-relevant information like URLs. The redirect script is shown in figure B.1. The modules are located in a windows called *external patient information*

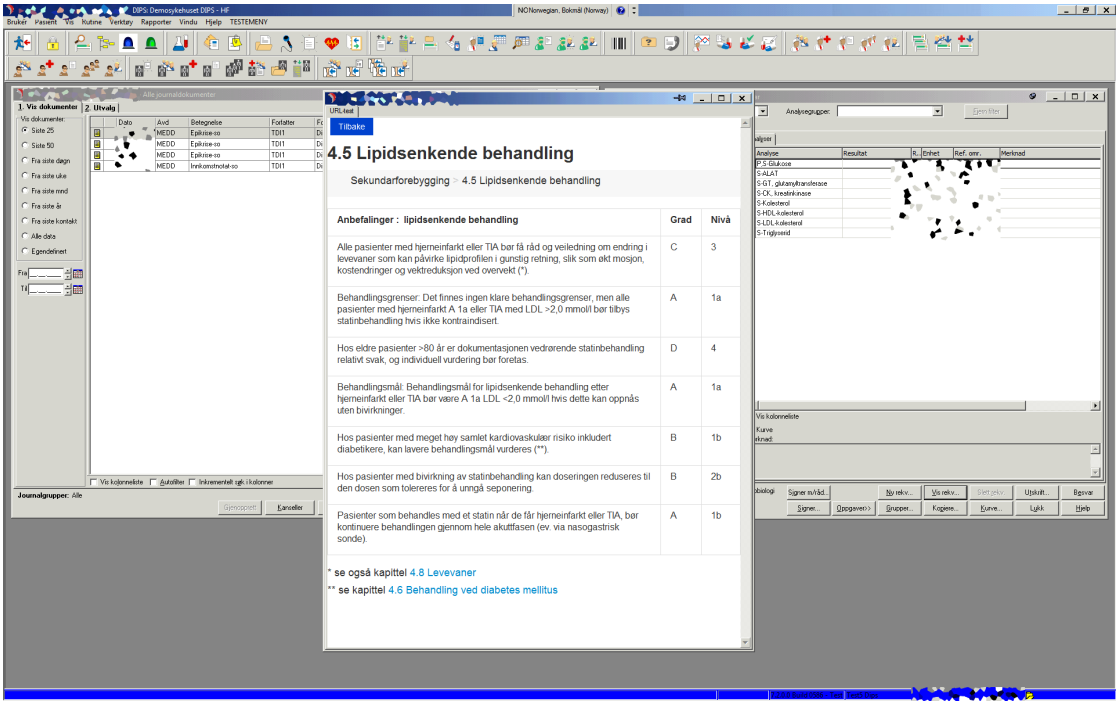


Figure 7.4: Guideline for cholesterol lowering treatment in GAS

7.5 Back End

Guidelines used in this projects are also useful for other projects under the EviCare umbrella. Helsebiblioteket.no lacks an official Application Programming Interface (API) for external use of guidelines, so guidelines are only accessable by browsing or searching through the website. Back end services at Helsebiblioteket.no stores guidelines in several different structures, and this project therefore only focuses on the structure of the stroke guidelines. These are extractable to XML, but contains irrelevant meta data and lack relevant meta data. This issue is discussed further in chapter 11.1.1: *API Access vs Scraping*. Timing is also essential here, it was quicker to access guidelines through the website, rather than through NOKC. So to access the guidelines automatically using the website, it is necessary to use scraping. Scraping is *the action of taking content from one web site for use un another*.
An offline Python script scrapes guidelines from Helsebiblioteket.no (link) using the Yahoo Query Language (see chapter 7.5.2), and translates the semi structured Hyper Text Markup Language (HTML) to structured files in XML format. These files include all recommendations from a selected set of guidelines (here:

the national stroke guidelines). Recommendations are given an ID in the format `recommendation#`, where # represents a number. This number is zero indexed, and the index starts at the first recommendation it finds, while incrementing through all recommendations in the selected guidelines.

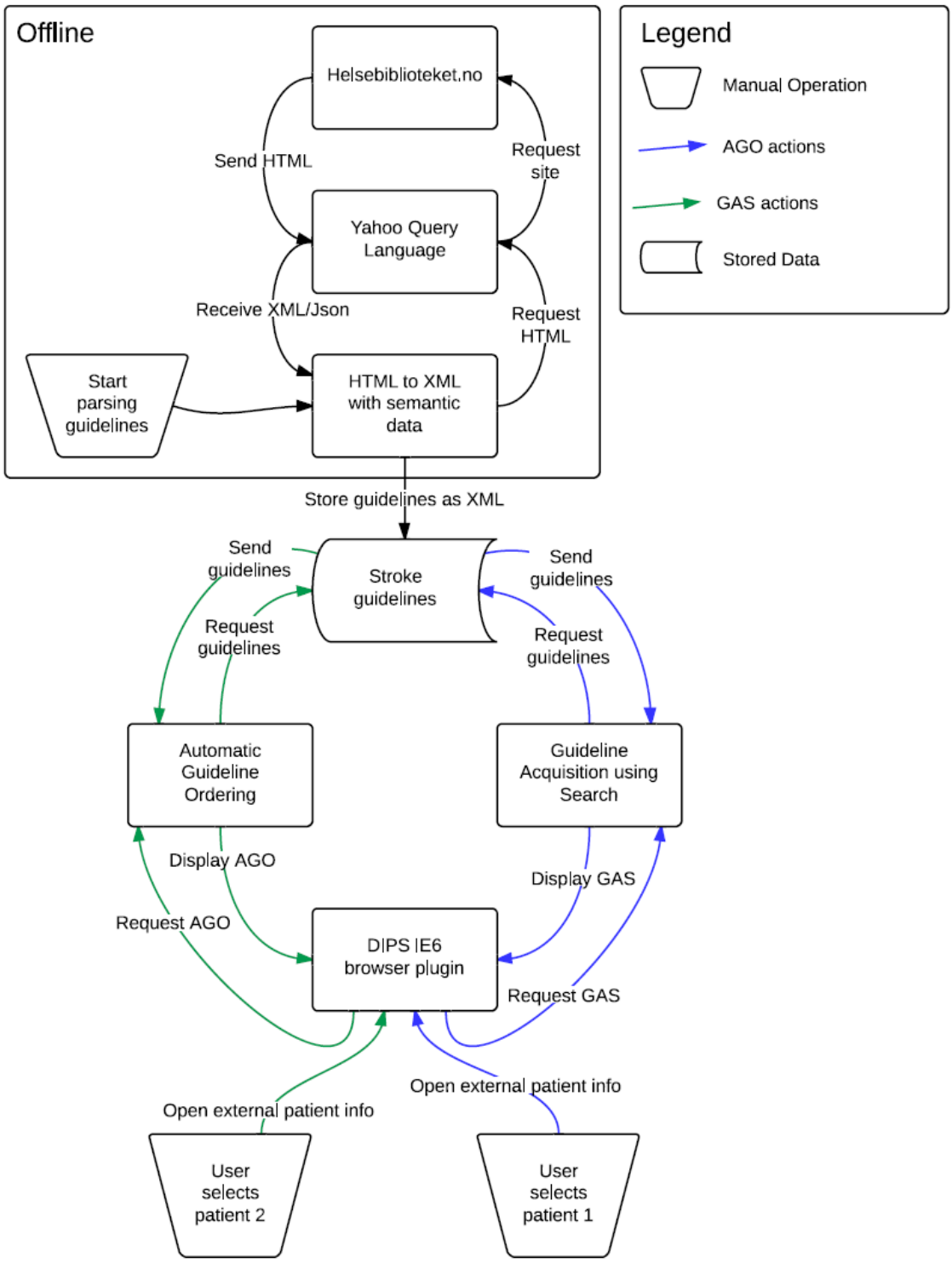


Figure 7.5: Overview over GAS, AGR and DIPS

7.5.1 XML structure

Results are stored in separate XML files for each guideline, where the hierarchy and metadata structure is as shown in figure B.2. It is a tree structure where the domain is the top node, and guidelines are child nodes. Recommendations are almost not altered from their original table structure, with the exception of adding identifying numbers in the attribute `recommendationID`.

7.5.2 Yahoo Query Language and XPath

Web scraping is done by using a popular third party tool created by Yahoo, called Yahoo Query Language (YQL). YQL is *an expressive SQL-like language that lets you query, filter, and join data across Web services* (Yahoo, 2013b). Yahoo provides a public API, for free, which simplifies web scraping for developers. The query is quite similar to queries in Structured Query Language (SQL), and supports XPath. XPath is a language designed for adressing parts of the XML structure, especially XML nodes. It was defined by the World Wide Web Consortium, and hit version 1.0 in 1999. (World Wide Web Consortium, 1999) Version 2.0 is the current version, but YQL currently supports version 1.0.
The script utilizes YQL for scraping the stroke guidelines (and other structured national guidelines, by applying some changes). It fetches table of contents, or an overview over all recommendations in the stroke guideline (or other chosen guideline, based on the url) from `http://www.helsebiblioteket.no/Retningslinjer/Hjerneslag/Innhold`. From there, it visits all links and sublinks still using YQL. Each link either contains a set of subchapters (example) or a set of recommendations example.
The script separates links containing recommendations from links containing subchapters using XPath. The XPath expression
`//*[@id='recommendations']/ancestor::div` searches for a node in the XML structure containing an `id` attribute where the value is 'recommendations'. It then selects the recommendation node, goes to the nearest ancester, and returns the content of this node.

7.5.3 Python

Python is a effective and powerful cross-platform programming language, inspired by other common programming languages like Java, Perl and Ruby. Main focus areas include readability, intuitivity, and syntax close to natural language.(Python, 2013) Python received its name from the Monty Python movies.

7.5.3.1 ElementTree

ElementTree is a default library for Python. This library enables developers to create XML node structure using simple functions. It is here used for creating the stroke guideline XML file, but may be altered to create similar files for other guidelines.

7.6 Front End

A JavaScript based script parses the XML file, with help of jQuery library. The actual web site content is structured using HTML and Cascading Style Sheets (CSS). Most of the CSS file is created using Twitters Bootstrap framework (Twitter, 2013), which makes the site much more appealing. Both jQuery and Bootstrap significally improves the development process, which in turn simplifies implementation of changes to the prototype. Figure 7.6 shows the initial design both with and without Bootstrap styling. Bootstrap also offers several features through JavaScript scripts, features such as expanding and collapsing elements. These features are used in AGR for expanding and collapsing guidelines, but were dropped later on in the process, which will be further explained in chapter 11: *Limitations*. For a more detailed view of the differences between Bootstrap styling and no styling, refer to appendix B.2.

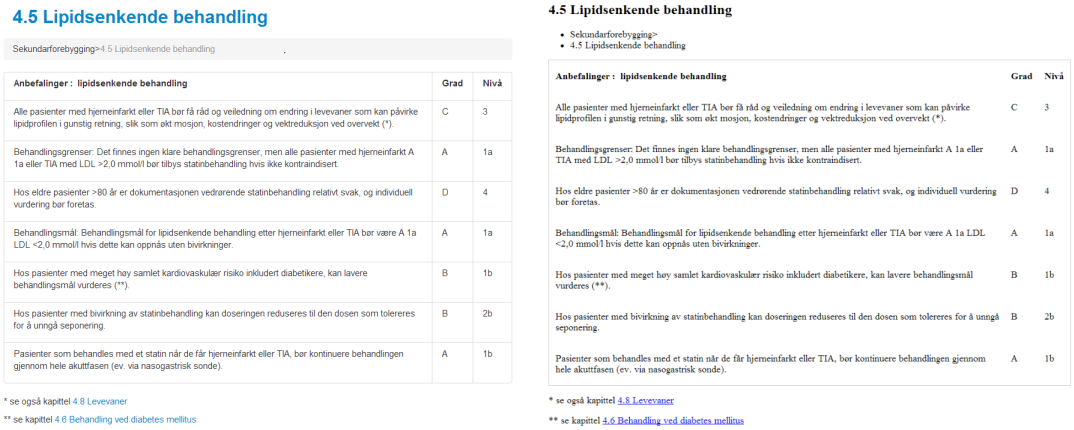


Figure 7.6: AGR modules initial design with and without Bootstrap styling

Chapter 8

Experiment Execution

This chapter describes the implementation and execution of the experiment design from chapter 6.

8.1 Recruiting Participants

The easiest way to recruit medical students at NTNU, is through their student organization Placebo (2013). Students were contacted per mail, through Placebos mailing lists for each class. The usability lab were booked from april 2nd to april 9th, calculating 1,5 hours per participant. 12 students participated in the experiment, this number were chosen on beforehand. Five to six participants is often enough to uncover common errors and flaws of any given prototype. But in order to ensure a significant amount of data for the analysis, 12 participants were selected.
A non-student were selected as the pilot participant. The purpose of this pilot were to uncover flaws with the experiment design, and to get a better approximation of time needed for each participant. The pilot participant were an graduate engineer working at NSEP, with knowledge about several clinical domains, and the use of clinical guidelines.

8.2 Observation

There were only one observer throughout the experiment, taking notes with time stamps from significant findings. Those findings included most actions directly related to searching for and finding guidelines, and evaluating their relevance. One example of significant action is where the user first enters the window for external patient information.

8.3 Usability Lab

All tests were performed in NTNU IDIs usability lab at NSEP. The lab contains a computer, the Tobii EyeTracker system, video cameras and microphones. Video cameras were directed at the participant, and recorded during task execution and interviews. Two microphones were directed at the participant and observer and recorded audio at the same time as the video camera.

8.3.1 Tobii EyeTracker and Tobii Studio

An eye tracker is a system that gathers data, data which calculates where on the computer screen (and effectively in the DIPS system) the participants are looking. This is done by illuminating the participants eyes and recording the reflection. Tobii EyeTracker is the eye tracker hardware used in this experiment, while Tobii Studio is the software suite for recording, altering and exporting eye tracker data. (Tobii, 2013)

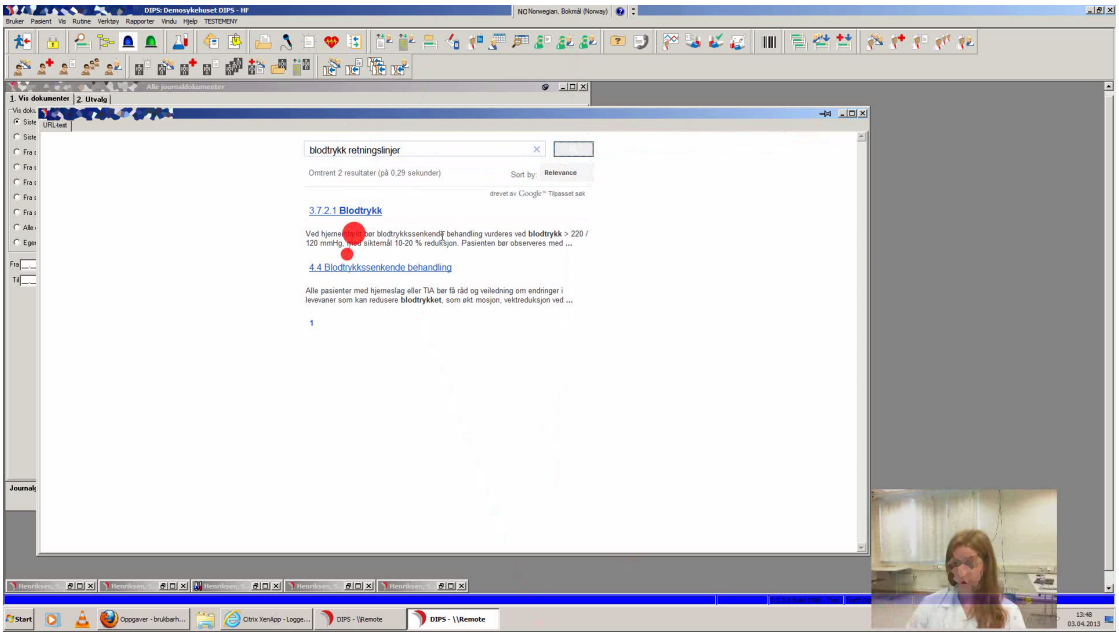


Figure 8.1: Screenshot of Tobii Studio

Eye tracking data is used by participants and observer/interviewer for evaluating the usability test and task execution. The red dots in figure 8.1 depicts eye movements, and simplifies the evaluation for participants. They clearly see what

their focus points were when using the system, which in turn facilitates reflection during interview sessions.

8.4 DIPS Test Hospital

DIPS helped set up a test version of the DIPS EHR system, a so called test hospital. This is the same system used for earlier experiments in EviCare, with some modifications. The test hospital contains a module/window called **Ekstern Pasientinformasjon**, meant to be used for external information about the patient. In this case, the window were used for fetching the guidelines.

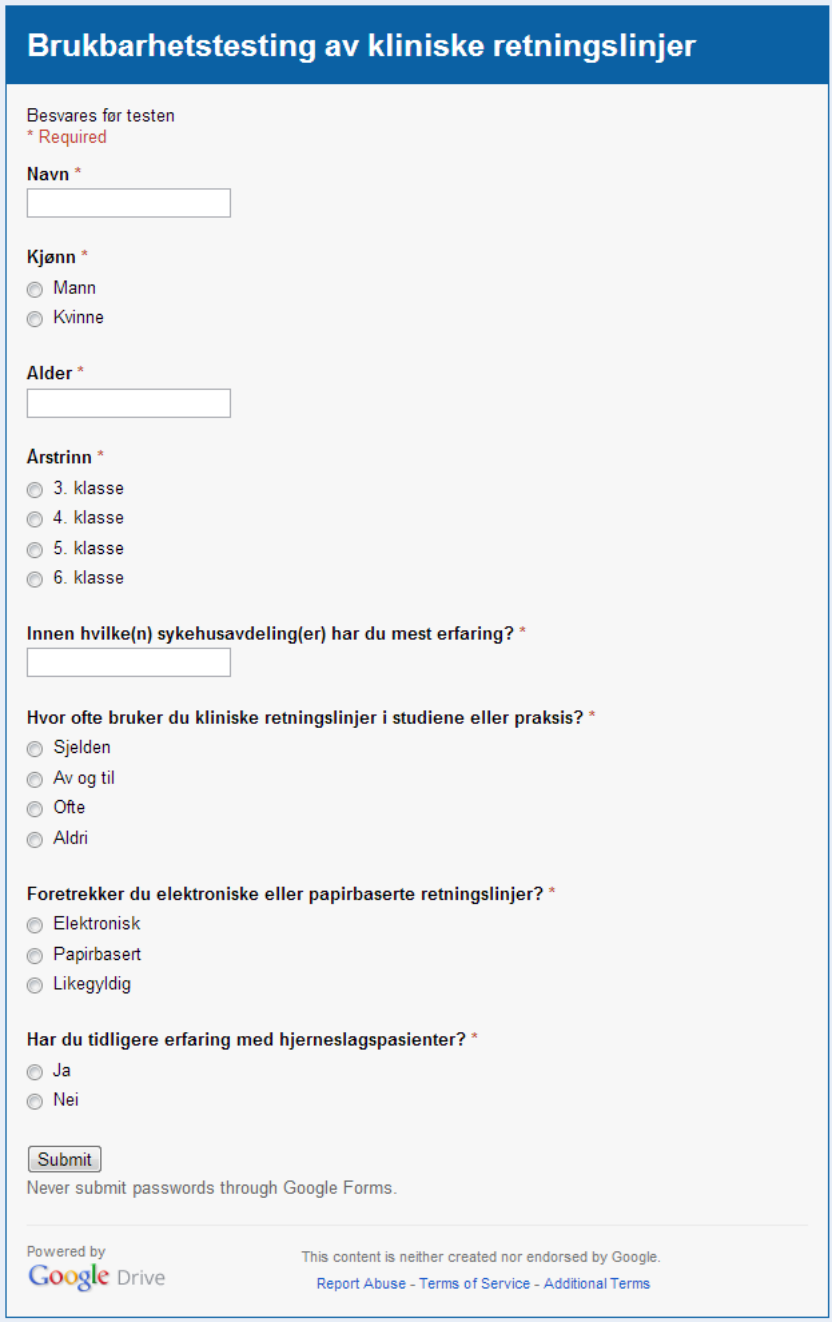
8.5 Google Drive

Presentation of surveys, forms and task list were done by using Google Drive. Google Drive is a collection of several tools for document creation, manipulation, publishing and sharing. Documents include spreadsheets, text, presentations and forms. Tools used in this experiment include the word processor called Google Docs and the tool for creating forms and surveys called Google Forms.

Google Docs is a collaborative tool for editing text documents. Experiment description, tasks, and other information were presented to the participants using Docs, through a web browser. All of the above were presented using only one Docs document, with links to surveys created in and hosted by Google Forms.

Google Forms is simply put a way to collect data through customizable online forms. All surveys in this experiment were created and published through Google Forms. Data collection and representation is automatic, which makes Google Forms fairly easy to use.

8.6 Preliminary Survey



Brukbarhetstesting av kliniske retningslinjer

Besvares før testen
* Required

Navn *

Kjønn *
- Mann
- Kvinne

Alder *

Årstrinn *
- 3. klasse
- 4. klasse
- 5. klasse
- 6. klasse

Innen hvilke(n) sykehusavdeling(er) har du mest erfaring? *

Hvor ofte bruker du kliniske retningslinjer i studiene eller praksis? *
- Sjelden
- Av og til
- Ofte
- Aldri

Foretrekker du elektroniske eller papirbaserte retningslinjer? *
- Elektronisk
- Papirbasert
- Likegyldig

Har du tidligere erfaring med hjerneslagspasienter? *
- Ja
- Nei

Submit

Never submit passwords through Google Forms.

Powered by Google Drive

This content is neither created nor endorsed by Google.
Report Abuse - Terms of Service - Additional Terms

Figure 8.2: Form answered by participants before usability tests

The preliminary survey was distributed via email prior to the usability test. Most participants answered the survey before they showed up at the usability lab, and some answered the survey at the lab, right before the test.

8.7 Preparation

Before the actual recording and first task, the participants were introduced to usability testing by the observer. They were made aware of their surroundings in the lab, and the fact that their actions during the test would be recorded (with answering of System Usability Scale (SUS) forms and preparation as exceptions). They were also explained that the test subject was the DIPS system, and not the participant. This helped ease their concerns about own competence and experience level. In addition to explaining the usability test method, some participants were also given the opportunity to refresh or gain knowledge about clinical guidelines. All introductory explanation were given both textually and orally.

8.8 Task Execution

After completing the introduction and EyeTracker calibration the recording started, and the task execution began. Participants were asked to do two sets of tasks, one for each guideline module and patient. The tasks in each task group were more or less identical, with minor differences. Tasks were given in norwegian, the same as the surrounding help text.
The first set of tasks focused on search:

1. Find and select patient "Henriksen, Stein"
2. Ensure that the selected patient is "Henriksen, Stein"
3. Find the related discharge notes
4. Find relevant guidelines using the window for external patient information
    - Open the window called "external patient information"
    - Search for relevant guidelines based on the discharge notes you read earlier

The second set of tasks focused on automated ranking:

1. Find and select patient "Havnut, Vidar"
2. Ensure that the selected patient is "Havnut, Vidar"

- The case history is the same as in task 1
- You are free to re-read the discharge notes

3. Find relevant guidelines using the window for external patient information

   - Open the window called "external patient information"
   - Find the most relevant guidelines based on the discharge notes you read earlier

In the end of each set of tasks, the participant were to evaluate guidelines and find the most relevant one, and then recommend further treatment.

8.9 System Usability Scale

After each usability test, the participants were asked to fill out a SUS form (Brooke, 1996). This form is basically a likert scale[1] (first invented by Likert (1932)), with scores from 0 to 5. A score of 0 represents strong agreement, while a score of 5 represents strong disagreement. Common SUS forms consist of 10 statements about a given system, where respondents tick of a number representing their level of AGReement. The data is then processed using a formula where the results give an indication of usability. SUS forms are useful in situations where multiple system or iterations of a system are compared with each other. Here, the AGR module were compared to the GAS module.
Both SUS forms were created and published in Google Forms, and responses were anonymously stored in Google Drive for analysis. There was no point in storing information about each responder, and anonymous surveys solves common ethical and legal dilemmas. An example of the SUS forms is located in appendix A.3.

[1]Definition of likert scale: *a scale used to represent people's attitudes to a topic*

8.10 Interview

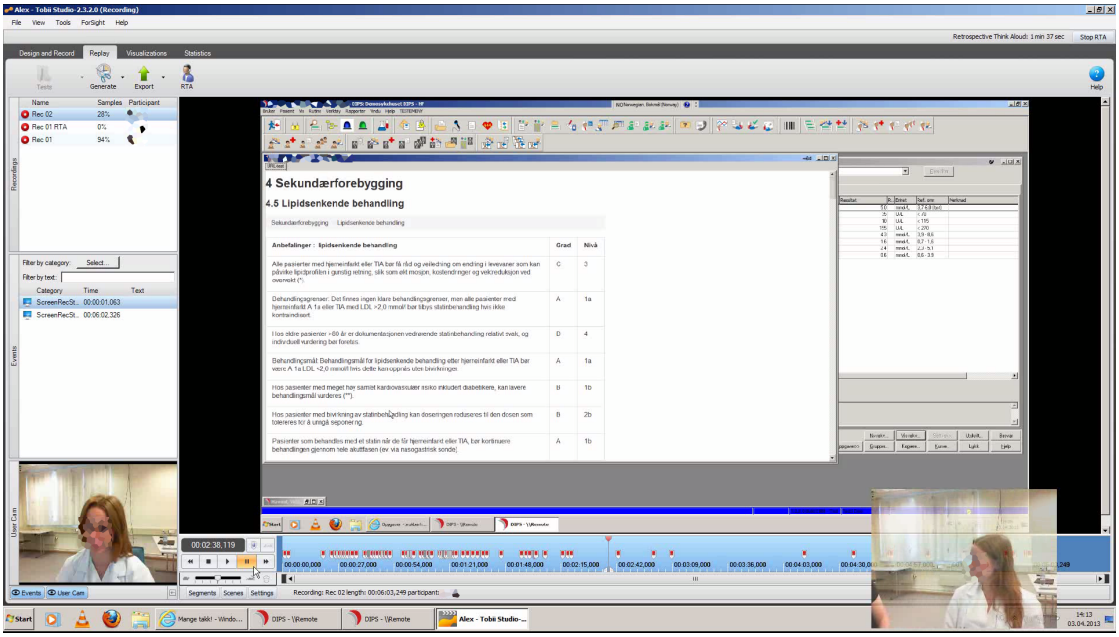


Figure 8.3: Screenshot showing Tobii Studio with DIPS and a video of the participant in the lower right corner

After answering the SUS forms, participants were asked several questions in an semi structured interview. The goal of the interview were to uncover more details and data around the usability test, in addition to give the participants an opportunity to elaborate on their SUS form answers.
To encourage reflection, participants watched a video of their own usability test, and answered questions related to a given task on the video. This combination of retrospective think aloud and semi structured interview made it easier for the participants to remember their actions, It also created a simple way to point out positive or negative elements in the design or workflow, by pausing, scrolling through the video and physically point at the screen while explaining. The video contained a small screen with a recording of the participant, a main screen showing the screen recording, and dots on the screen depicting eye movements. There were two sessions of retrospective think aloud and interviews, one after each set of tasks.
Questions varied throughout the experiment, tailored for each individual situation and participant. Although questions were adapted and changed, some main themes remained despite of the specific context. Recurring themes included:

- recommendation grading

- user satisfaction in both modules
- credibility
- next step of action/most relevant guideline
- suggestions for improvements
- information amount
- preferred method for guideline fetching

Chapter 9

Results

This chapter presents the results from the usability tests, interviews, surveys, and system usability scales. Only results relevant to the research questions are included, with other significant findings.

9.1 Preliminary Survey

As mentioned in chapter 6.7, the participants received a survey before the experiment execution. This subchapter comprises the resulting data from those surveys. Some identifying data are left out, in order to anonymize participants.
When asked about their preferred method of guideline browsing, ten participants preferred electronic guidelines, two were indifferent and no one preferred browsing through paper based guidelines (meaning books, printouts etc.).

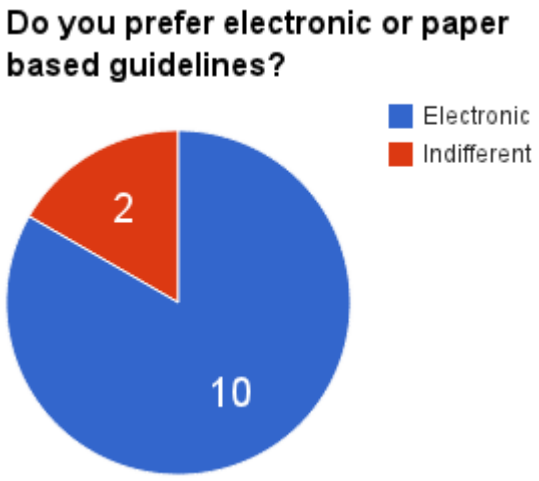


Figure 9.1: Chart showing what method of guideline fetching participants prefer

Experience and knowledge level was an important factor in this experiment, and the easiest way to map a students basic knowledge level is to determine their class year. Both experience and knowledge vary from participant to participant, but class year gives a satisfying approximation. So one question in the survey simply asked them to provide their class year. The majority of higher grade students are less available than lower grade students, due to obligatory clinical practice. Therefore, the largest subgroup in the participant group consisted of six third grade students. The other half of the participant group consisted of four 4th grade students, and two 5th grade students.

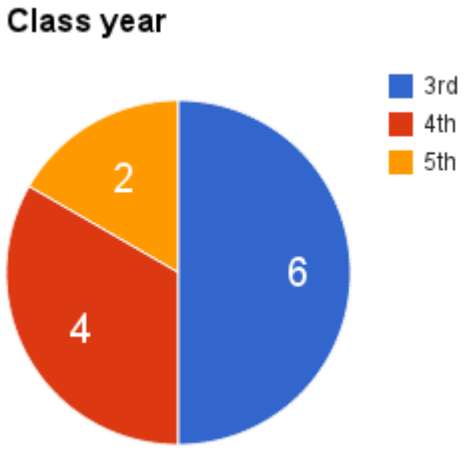


Figure 9.2: Chart showing participants class year

Another way to measure their experience level is to map their knowledge and experience with certain elements from the test like stroke patients and national clinical guidelines. Data from the survey show that half of the participants use guidelines in any form "now and then", while two seldom access guidelines. Only one participant had no prior experience with guidelines. The remaining three participants answered that they often benefit from guidelines.

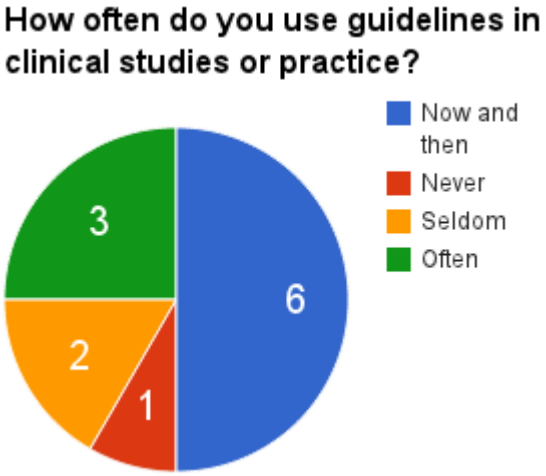


Figure 9.3: Chart showing how often participants use clinical guidelines

When it comes to stroke patients, almost everyone (10 of 12) answered that they have previous experience with that domain.

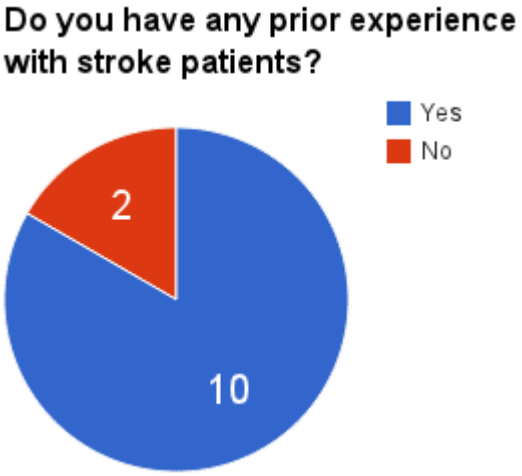


Figure 9.4: Chart showing participants prior experience with stroke patients

In addition to mapping their background with stroke patients, they also listed other domains they have practiced. The table in figure 9.5 lists the submitted domains, and the number of participants which have prior experience with each domain.

Clinical Domain	Number of Participants
Neurology	4
Psychiatry	3
Stroke	2
Gastro	2
Internal Medicine	1
Intensive Care	1
Skin	1
Geriatry	1
Heart	1
Gynecology	1
Orthopedy	1

Figure 9.5: Table of clinical domains practiced by the participants

9.2 Usability test

Many of the participants understood and adopted the think aloud method during the usability testing and task execution. Some did however fail to think aloud during most of the test, and some used it rather sporadic.

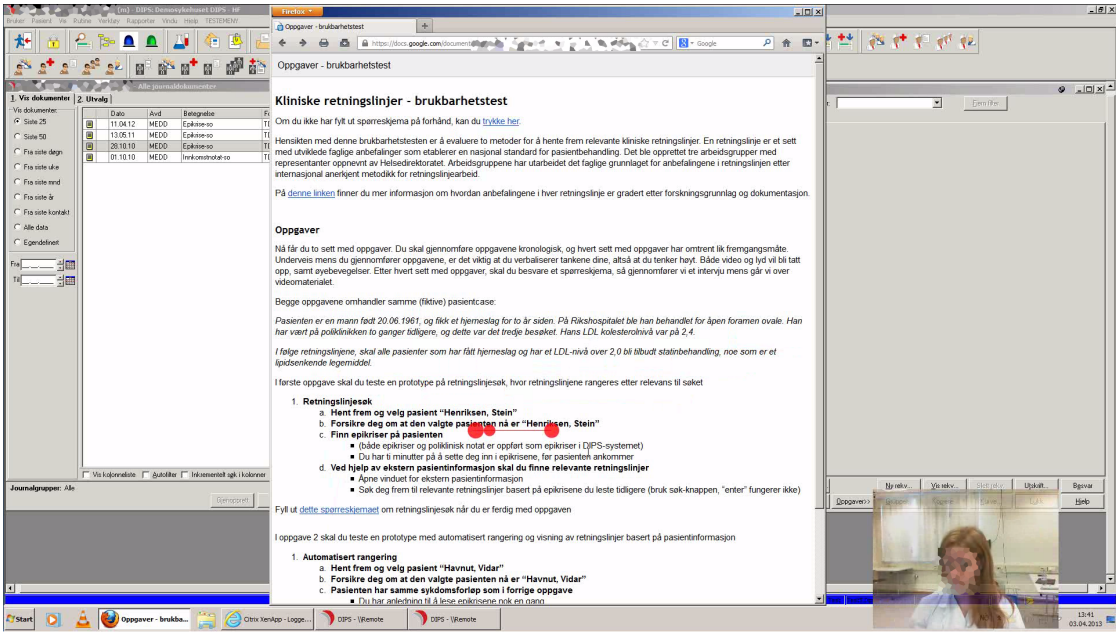


Figure 9.6: Screenshot of DIPS and tasks in a browser window

As mentioned earlier, the participants had the opportunity to revisit and re-read the discharge notes at any time during the usability test. Most read them and got an overview during the first task, and never revisited them. A portion of the participants revisited them during task 2 and explained that they had to refresh their memory.

9.2.1 Guideline Access using Search

Participants used the search functionality in GAS as they would use the Google search engine. A common approach were to enter broad search terms in the beginning, evaluate the results, and then narrow down the search. They did this until they found what they were looking for, or until they found themselves at a dead end or loop. Some struggled to get the preferred results, although they used case relevant search terms such as **hjerneslag (stroke)**, **kolesterol (cholesterol)**. Eight found the guidelines they were looking for. That meant finding the guideline on **cholesterol lowering treatment** (or **lipidsenkende behandling**)
A fraction of the participants still continued the search for other relevant guidelines after acquiring the most relevant guideline. They justified their choice by saying they wanted to ensure that no guidelines with high relevance were overlooked.
The following table display search terms used by the participants, and total frequency across all participants. Stroke, statins and LDL (low density lipoproteins)

amongst other words mentioned in the patient case were the most frequent search terms.

Search Term	Frequency
LDL	9
Hjerneslag	8
Statin	7
Platehemming	4
Patent Foramen Ovale (PFO)	4
Slag	4
Kolesterol	2
Hjerneinfarkt	2
Lipidsenkende behandling	1
Blodtrykksbehandling	1
Sekundær profylakse	1
Hjerte	1

Figure 9.7: Table of search terms and total frequency

9.2.2 Automatic Guideline Ranking

As mentioned in the previous subchapter, participants learned how to use certain DIPS components in task 1 (with GAS), and as a consequence, task 2 execution were more rapid. Participants expected the GAS module when opening the window for external patient information. They were therefore suprised to see guidelines appear without any searching or other navigation. Some described the design of the AGR module as almost identical to the individual results from task 1. A few of the participants saw the lack of navigation in order to access guidelines as a positive feature. These statements often came from participants which had struggled with task 1 and the GAS module, not to mention the DIPS user interface (but this is not a part of the usability evaluation).

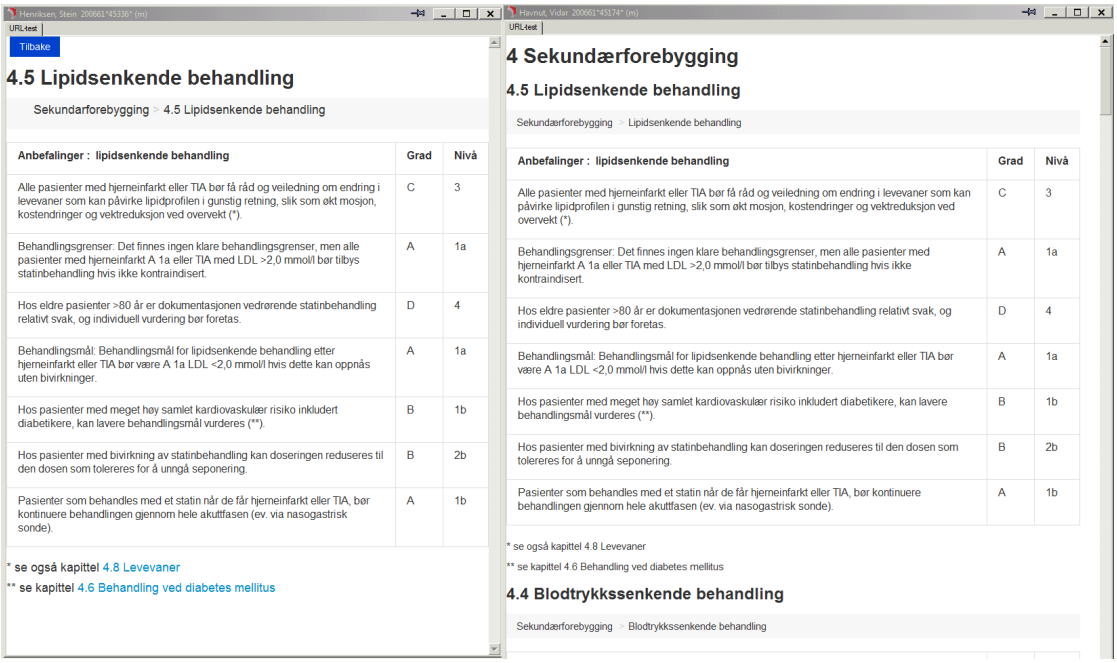


Figure 9.8: Comparison of GAS and AGR

As a result of this, some failed to see that the AGR module featured a list of several guidelines. That problem only occurred to a minority of the participants, but feedback from the interview suggested that others might experience the problem.

9.2.3 System Usability Scale

Participants were asked to consider the whole DIPS system when filling out the SUS form. That resulted in quite low scores, as the many of the participants viewed the DIPS system as difficult to learn. Participants lack previous experience with this system, and expressed frustration and pointed at issues directly related to the DIPS system design as the source of their difficulties. But the main focus from this test is to uncover differences between the GAS and AGR modules, and to see how they both affect and help the workflow of clinicians. It is therefore interesting only to compare SUS scores between those two modules, and compare those to possible future iterations. Both systems received very low scores, and the lowest score were given to the AGR module, with a calculated score of 24. In comparison, the GAS module received almost the double of that, with a calculated score of 42. See chapter 8.9 for information on how the score was calculated. The following diagrams present results from both SUS forms:

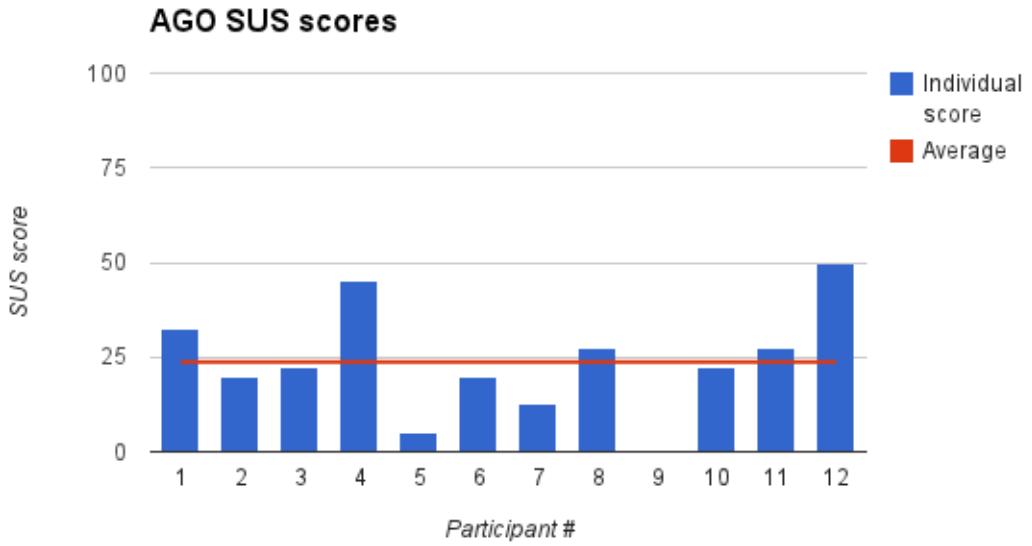


Figure 9.9: SUS scores from the AGR module

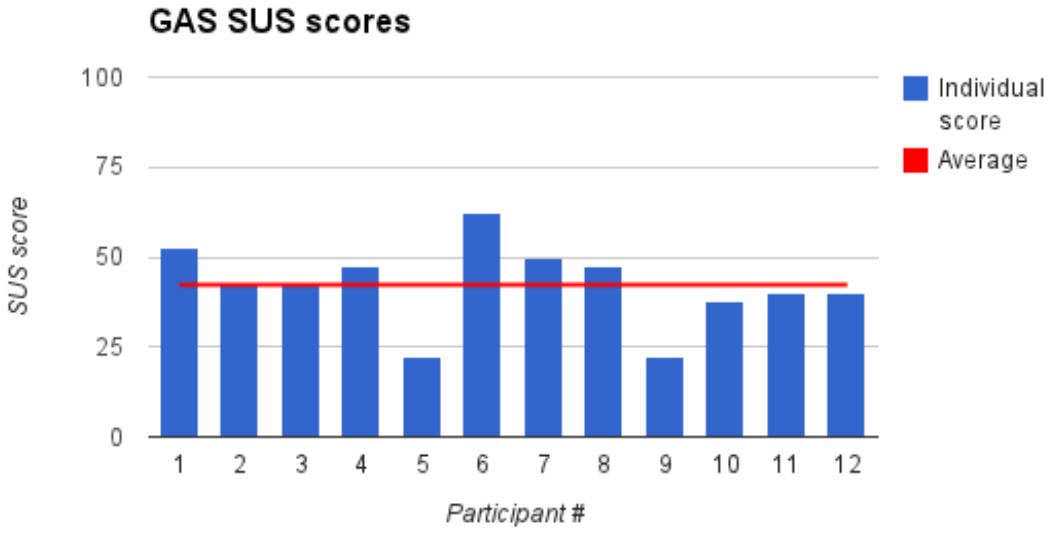


Figure 9.10: SUS scores from the GAS module

9.2.4 Interview

The interview were conducted right after the participants filled out the SUS form, and therefore made them elaborate and reflect around their answers. Although some also mentioned design flaws with the DIPS system, interview questions were mainly directed at the integrated modules.
The main goal of the interview were to fill in holes from the usability test, especially where the participants failed to use the think aloud method throughout the task execution. Questions were also composed with the overall experiment goal in mind. Those include user satisfaction and changes in workflow, which best can be measured both individually per module and by comparing them.
One of the main issues by using these guidelines with recommendations, is the credibility. To what degree does the clinicians and other users trust the data, and what could be made to improve the overall credibility? This chapter presents feedback regarding this issue, then feedback on the preferred way of obtaining guidelines. Arguments for each module are also presented, while the two last subchapters mention general feedback for each one.

9.2.4.1 Credibility

One experienced participant were sceptical of low graded recommendations, since it is known that those recommendations are most likely controversial. The same participant have also experienced that different hospitals use different guidelines, and wondered if the presented guidelines actually represent the guidelines used at St. Olavs hospital. In order to compensate for this uncertainty, some participants wanted more background information around grading and research information directly accessable in the GAS and AGR modules. Other participants saw no credibility issues with the national clinical guidelines.

9.2.4.2 Recommendations

Almost half of the participant group (5 of 12) wished for the recommendations to be sorted internally in each guideline. Suggestions ranged from sorting by chronological order to sorting by grade and level of research background and documentation.

Preferred method of guideline fetching

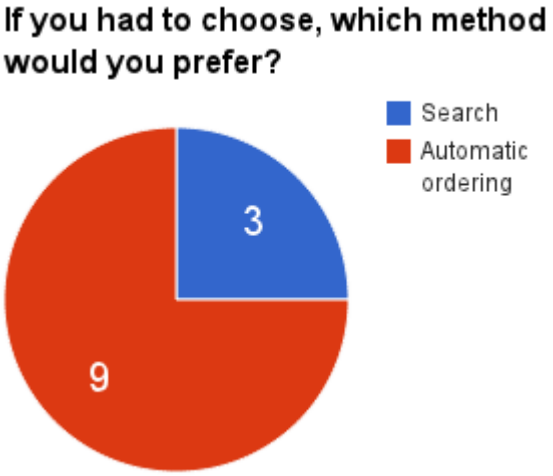


Figure 9.11: Chart showing which module the participants preferred

AGR were a clear winner when participants were asked to choose between modules. 75% of the participants dubbed AGR as their preferred method of guideline fetching. Those results goes against all results from the SUS forms. The inconsistency could be explained in several ways, which will be explored further in chapter 11.
Multiple arguments supporting AGR were phrased during the interview. Results show that although the interviews were conducted individually, some arguments were repeated throughout the experiment. Over half of the participant group listed the argument that AGR includes guidelines they had not thought of by themselves. Several participants also pointed out that AGR improved the workflow significantly by eliminating the need to alternate between a result page and a page for browsing guidelines. GAS' functionality worked in that manner, which helped the participant view the workflow in AGR as quicker.
AGR lists several guidelines in a one page design, giving clinicians a full list of multiple relevant guidelines. The list provides an abundance of alternatives to choose from, and participants identified this feature as critical for usability. After working with professional clinicians, participants noticed that some clinicians often avoid tools that disrupts or slow down their workflow.
AGR were also described as to being easier to use for those inexperienced with clinical decision support systems. One participant also noted that AGR "feels safer".
Arguments for GAS were fewer in number and none were mentioned by more than one participant. There seemed to be a pattern where more experienced students preferred GAS, while inexperienced students chose AGR. Those who chose GAS as their preferred way of obtaining guidelines, pointed out user control as one of

the main arguments for GAS. There is situations where clinicians want to look up guidelines which are irrelevant to the selected patient case.
While almost every participant deemed multiple guidelines as relevant, most selected a distinct guideline as the most relevant one. Few participants changed their views on that matter from task 1 using GAS to task 2 using AGR. AGR mostly confirmed the views of each participants by ranking their chosen guideline as the topmost relevant one. Some participants did however notice that their selected guideline were in fact ranked lower than the most relevant one in AGR.
Nine out of 12 participants chose the guideline for cholesterol lowering treatment as the most relevant guideline when using AGR, that is, the same guideline that were predefined as the most relevant one. Almost as many chose the same guideline to be the recommended way of treating the patient when using GAS. Figure 9.2.4.2 shows what guidelines the participants chose as the right approach when asked to give recommendations for further treatment.

Participant #	After using GAS	After using AGO
1	- Cardiac infarction and transient ischemic attack (TIA) - Patent foramen ovale	Cholesterol lowering treatment
2	Blood pressure lowering treatment	Blood pressure lowering treatment, then cholesterol lowering treatment
3	Cholesterol lowering treatment	Cholesterol lowering treatment
4	Cholesterol lowering treatment	Patent foramen ovale
5	Cholesterol lowering treatment	Cholesterol lowering treatment
6	Cholesterol lowering treatment	Cholesterol lowering treatment
7	Cholesterol lowering treatment	Cholesterol lowering treatment
8	Cholesterol lowering treatment and stroke, but unable to find the right guidelines	Cholesterol lowering treatment
9	Patent foramen ovale, to ensure that previous treatments is according to the guidelines, and that there's no further treatment here	
10	Cholesterol lowering treatment	Cholesterol lowering treatment
11	Cholesterol lowering treatment	Cholesterol lowering treatment
12	Patent foramen ovale	Cholesterol lowering treatment

Figure 9.12: Table showing recommendations for further treatment

9.2.4.3 GAS feedback

In addition to arguments for either AGR or GAS, the experiment resulted in more general negative and positive feedback. This subchapter presents feedback gained by the GAS module. Participants also suggested multiple improvements to solve the issues they uncovered. These are presented at the end of this subchapter. Almost half of the group experienced some degree of uncertainty when using GAS for the first time. Participants were unsure as to which search terms to use for finding what they want in an effective manner. After finding the right search terms, the participants described the GAS module as "logical". Three participants did how-

ever discover that some synonymous terms gained different results. Another group of three participants requested more hits in the results list, possibly to include hits from synonymous search queries, almost like what Google has implemented in their main search engine.
Both the results list and the actual guideline pages were described as short and concise by some participants. But some also felt that time were wasted when skimming through several irrelevant guidelines in the results. Those who started searching broader and refining their queries based on previous queries, felt that they eventually found what they were looking for. So the group were split in the question of how effective GAS was in providing the right guidelines. One participant noticed that the guidelines were numbered and wanted to browse through the parent categories. In addition to feedback on the results, one participant provided feedback on the idea of an integrated GAS module in DIPS. The fact that clinicians only have to focus on one application, without having to open an external browser window were brought up as a positive and welcoming change.

Improvements
As mentioned earlier, participants provided several suggestions for improvements to the GAS module. Improvements which may increase user satisfaction and system credibility. Those improvements are largely based on issues discovered by participants, and meant to solves these issues.
One third of the participants wanted a list of the main chapters or categories where the headings are clickable. When clicked on, these reveal subchapters related to the selected chapter. Some wished for features implemented in other web pages, for instance automatic search (Google) and links to more information about a subject (Wikis like Wikipedia). Automatic search provides a solution in which the search engine suggests search terms as the user writes them down. Google Search are also able to automatically search for these terms before the user has entered the complete query. Participants wanted to copy other features from established search engines like advanced search. Per their suggestions, advanced search could include searching by ground state (like "stroke"), with category or chapter filters. In order to increase credibility, the results must be ranked by the actual relevance to the search term. Some participants did however suggest an improvement where the clinician define the sorting method.

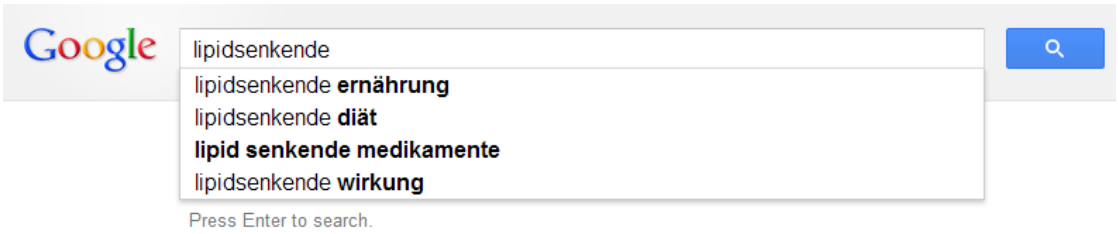


Figure 9.13: Google Autocomplete in action

Guidelines could link to more external information, or to other guidelines, as suggested by two participants. Links may also benefit the result page, some pointed out. The front page of GAS could have included most frequently used search terms. When clicked, the module produces results based on the selected term.

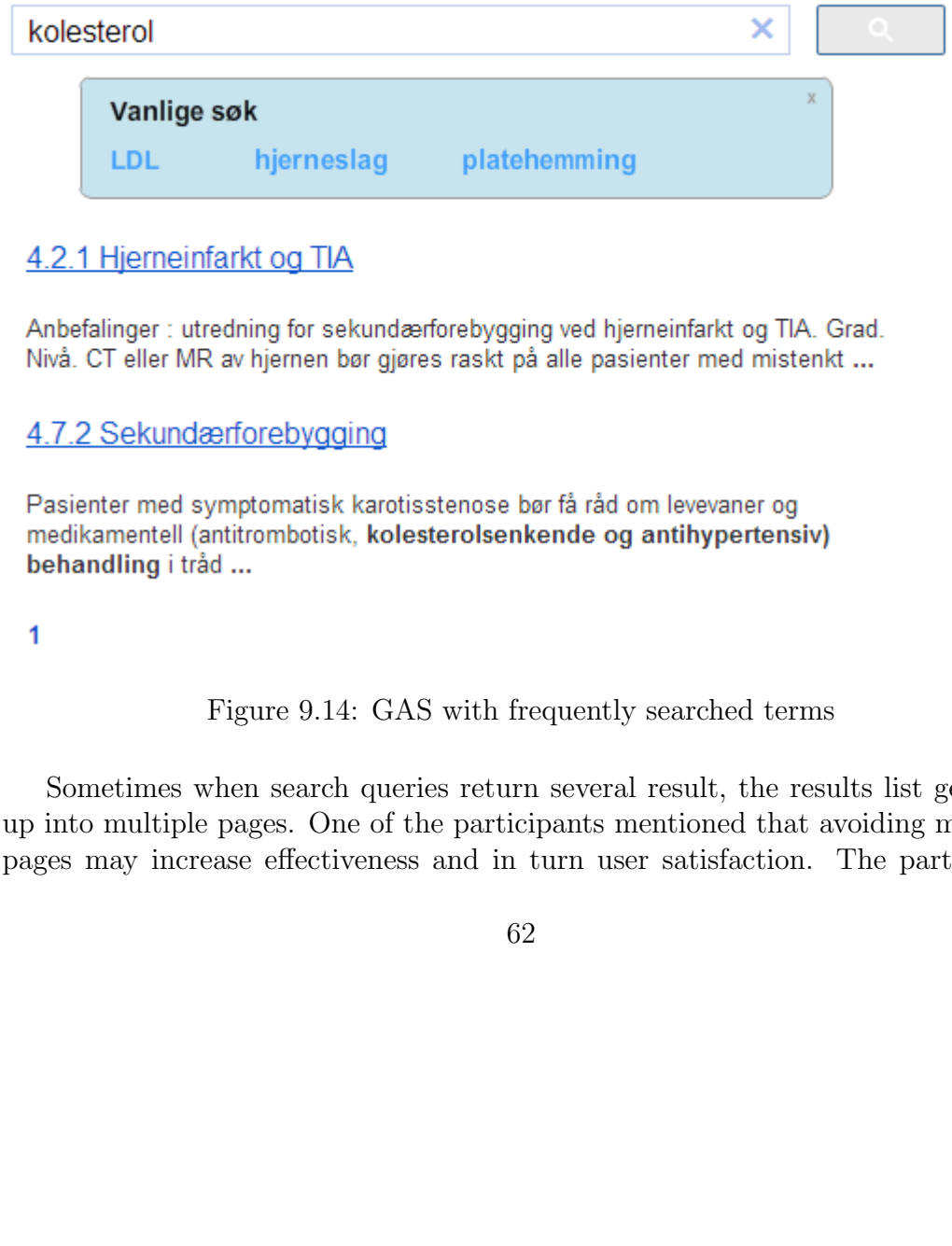


Figure 9.14: GAS with frequently searched terms

Sometimes when search queries return several result, the results list get split up into multiple pages. One of the participants mentioned that avoiding multiple pages may increase effectiveness and in turn user satisfaction. The participant

suggest a single page design, where all results are listed in one page, similar to popular sites as Facebook and Twitter. Then, the results loads as the user scroll down the list, demanding less user interaction. In addition to that change, a participant also suggested to make the content of each result element clearer.

9.2.4.4 AGR feedback

The AGR module too received both positive and negative feedback. Feedback were more harmonized in this interview, participants largely agreed on the general feedback. Almost half of the participant group perceived the AGR module as easy to use, but some were confused after using GAS. They questioned the way guidelines were ranked, since semi-relevant guidelines were placed far down the list. Some also confused AGR with the guideline page from the GAS module. The list looks as though it only contains one guideline, while others are located out of sight, further down the list. All participants saw this after some use, with one exception.
Some participants stated that they would have wanted more control over the process of selecting guidelines for the AGR results list and defining patient case relevance. Critical thinking were emphasized by one participant, since the system provided no guarantees as to whether or not all relevant guidelines got included.
The ranking system lost some credibility when participants discovered relevant guidelines ranked below irrelevant guidelines. One participant mentioned that clinicians probably will have to use the system multiple times to confirm that the ranking actually works as intended. Others noted the fact that cholesterol lowering treatment were at the top as a positive sign. Participants also pointed out that they scrolled through the irrelevant guidelines effortlessly. The importance of each guideline is individually decided by the clinician using AGR. Clinicians may wish to refresh the memory, as some of the participants mentioned. As the GAS feedback, feedback here suggested that the AGR module could improve usability by adding dynamic elements. AGR were described as being static, and by making it more dynamic, the clinicians get more involved.

Information amount
There were mixed opinions on the information amount presented by AGR. Some stated that too much is better than too little, as absent information would have to be obtained through other means. This meaning external applications like a web browser, or paper based sources.

Improvements
Participants suggested several ways of improving the AGR module as well, and

were fairly homogenous. 50% wanted to get an overview over the guideline headlines in a centralized place. They suggested a simple list or menu of the headlines represented as links. These links would bring the user right to where the selected guideline is located. Some also suggested other navigational upgrades. The most popular suggestion (excluding the list/menu suggestion) were to integrate AGR and GAS into one great module. Participants wished for the AGR module to do the dirty work at first, by presenting relevant guidelines. After that, they wanted the opportunity to search for other relevant guidelines, or search in the list of relevant guidelines. The latter could be resolved by enabling page search with ctrl+f, as this is the standard shortcut keys for searching a document of any kind (on PCs).

In order to increase AGRs credibility, participants suggested that the system would uncover what data the ranking were based on. Some pointed out the guideline for patent foramen ovale as being ranked too low.

Others wished for greater integration of AGR in DIPS, while improving the workflow. One participant suggested to have the AGR window available as soon as the clinicians open a discharge note. This solution could interrupt the clinicians workflow though, as stated by the participant, so the solution must focus on an non-interrupting design (eg. always having the window for external patient information visible). As mentioned earlier, many participants failed to see that the AGR results list were in fact a list of guidelines, and not just one relevant guideline. The addition of a menu or mini list could solve this problem, still one participant also suggested to add an indicator at the bottom of the visible page. This indicator would signalize that there is more guidelines in the list, and that they become visible by scrolling. Another solution to the same problem could have been expanding and collapsing guidelines, as suggested by another participant.

Chapter 10

Discussion

This chapter discusses experiment results, sheds light on general limitations with the experiment and modules, and outline future work with similar experiments and integrated CDSS modules.
Chapter 9: *Results* mentions how participants struggled with finding the window for external patient information. This could be a case of bad usability, which are not measured in this thesis, but mentioned in other theses. (Zheng Wang, 2012; Terje Røsand, 2012) Bad usability in the DIPS system will affect results when measuring AGR and GAS usability. Chapter 11 elaborates on this issue, among other limitations. First, it is important to discuss general feedback on both modules, and how they affect user satisfaction.

10.1 Guideline Access using Search

Usability testing of GAS were hindered to some degree because of a sub-par search engine for clinical terms. The standard Google Search engine is not trained for clinical terms, and this created some problems as explained in chapter 9 (and will be discussed further in chapter 11). A custom search engine for this domain should be built from scratch and trained for clinical terms, especially synonyms. Acronyms and abbreviations should also be considered, but as participants mentioned, these may be different from hospital to hospital. **LDL** were the most frequently used search term, alongside **hjerneslag (stroke)**. This means that most participants searched for the acronym, and no participants tried searching for the full length term (**low density lipoproteins**) as shown in figure 9.2.1.
Results also show that participants did a cycle of search, evaluation, refining search query, and search again. (See figure 10.1) They went like this until they had found the guideline they were looking for, or until they gave up. The latter was the case with one participant, as mentioned, and the situation made the participant un-

sure. This shows how underlying algorithms and functionality emerges and affects usability. Some got accustomed to the situation, and continued this cycle even after they had found what they viewed as the most relevant guideline based on the patients EHR. Participants explained it by stating that they wanted to ensure that they had found all guidelines relevant to the patients current status. So by observation and interview answers, one could argue that participants viewed the search as unreliable, since they thought it had omitted relevant results.

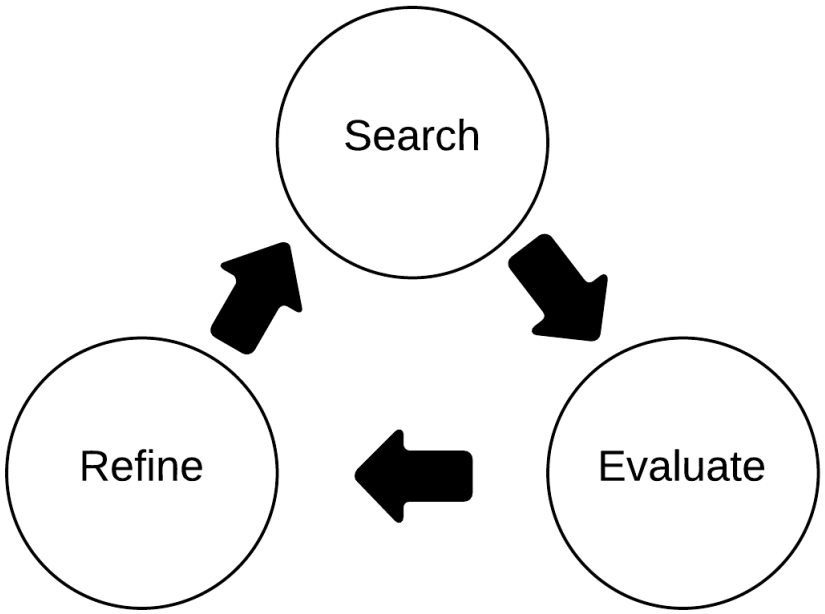


Figure 10.1: Search cycle in GAS

In addition to these findings, interviews discovered that some participants wanted and missed the **advanced search** feature from other familiar search engines. Usability studies show that consistency is key, this also means that GAS should optimally look and feel like comparable systems. As Jakob Nielsen puts it;

> Consistency is one of the most powerful usability principles: when things always behave the same, users don't have to worry about what will happen. Instead, they know what will happen based on earlier experience. Every time you release an apple over Sir Isaac Newton, it will drop on his head. That's good.
>
> The more users' expectations prove right, the more they will feel in control of the system and the more they will like it. And the more the system breaks users' expectations, the more they will feel insecure.

> Oops, maybe if I let go of this apple, it will turn into a tomato and jump a mile into the sky. (Nielsen, 2011)

Google Search is an obvious system to compare any search interface with, as well as search interfaces in PubMed and UpToDate. Both UpToDate and Google have advanced search, but more subtly than before. Google Search makes advanced search available only after the initial search. UpToDate have advanced search from the start, by allowing topic search, but makes the selected topic clear and visible to the users. It also selects "All topics" as the default option, which makes selecting another topic an active user operation. This level of user interaction helps create an awareness of context for the user. UpToDate also displays the topic in the results page, still providing the user a awareness of context. (Nielsen, 2001) These are some examples of interface features that could benefit GAS which also build upon user feedback. More on suggested future work in chapter 13.1

10.2 Automatic Guideline Ranking

Responses show that participants preferred the AGR module as a tool for acquiring clinical guidelines, over GAS. It does however have some properties with potential for further improvement. First and foremost; the list. AGR is basically a simple list when it comes to what the clinicians see and interact with. Therefore, this list is as important as any other background sorting and ranking algorithm. As mentioned previously, lack of awareness is a recurring issue in CDSSs (see chapter 2.3). While AGR greatly improves upon this issue compared to GAS, it still lacks some features in order to be a system that benefits clinicians. Clinicians are met with a guideline dubbed as the most relevant one at the top, with little or no apparent signs of the preceding less-relevant guidelines. Feedback from participants in the experiment suggest that this could lead to lack of awareness. Future designs would need to consider this, see chapter 13.

After overcoming the hurdle of actually scrolling down and uncover the additional guidelines, most participants had a better experience than with GAS. Most participants mentioned that they found guidelines which they did not consider or remembered to search for when using GAS. This clearly shows that AGR makes more guidelines available to the clinicians, and these are guidelines which clinicians fail to remember. Of course, this may only be applicable to this group of participants; medicine students. Professional clinicians would most likely remember numerous guidelines from years of experience. Further experiments must be executed with a variety of clinicians in order to conclude that a module like AGR will increase awareness and user satisfaction, improve clinical workflows and therefore improve patient care.

Another significant finding is that participants pointed out screen alternation as

an issue with GAS. Screen alteration in this case means that clinicians would have to alternate between the search interface and different guidelines. When GAS are switched out with AGR, user satisfaction seemed to increase as a result of less *back and forth*. Inexperienced clinicians prefer solutions that requires less user interaction, namely situations where the clinician must evaluate guideline relevance with less assistance from the CDSS than in alternative solutions (like AGR). AGR features a lower threshold for acquiring guidelines, since the most relevant guidelines are presented from the beginning. These differences may of course level themselves out if and when both modules get improvements.

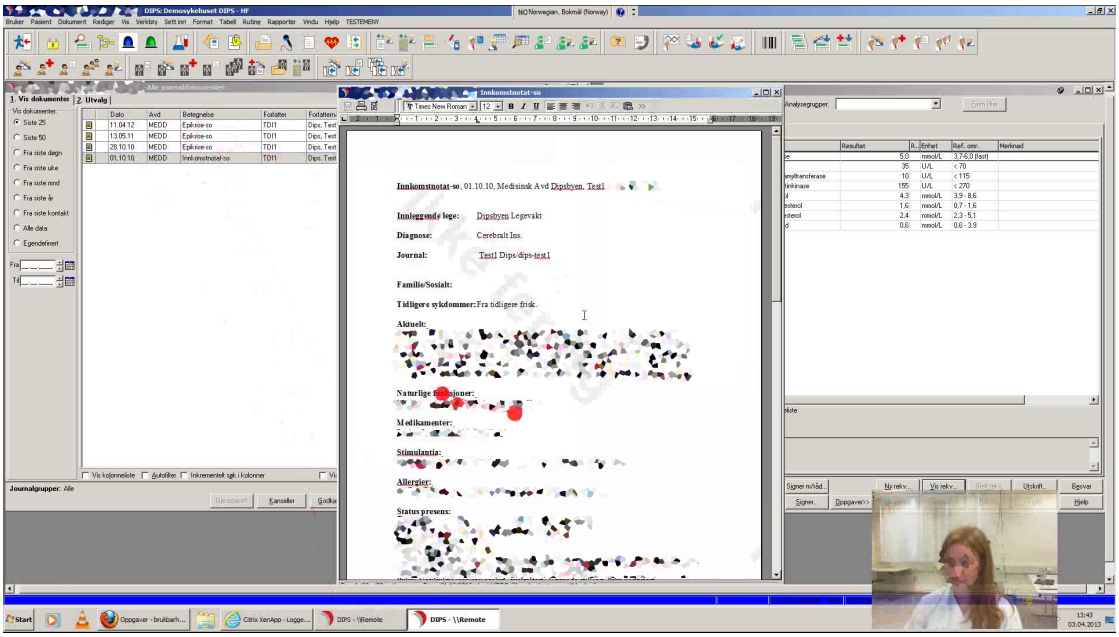


Figure 10.2: Screenshot from Tobii Studio with discharge notes in DIPS

10.3 Overall credibility

Both guideline and overall crediblity are significant for user satisfaction, and low credibility may create a situation where experienced clinicians refuse to take in new and updated clinicial recommendations. Individual recommendations are graded by work groups formed by the Norwegian Directorate of Health, and these work groups comprise of several clinicians from different hospitals and medical offices from different parts of Norway. Sources for the stroke guidelines are well documented and methodical strong, and picked out by said working groups. The grading system is therefore considered as a credible way of evaluating recommen-

dations. Credible sources and trustworthy recommendations leads to agreement from clinicians, and no fallbacks to previous practice.
While the basis for this CDSS is well documented and tested, participants had to lean on their own experience and knowledge in order to be certain. Both AGR and GAS received feedback which indicates that none of them accentuate the research which recommendation grading is based on. Recommendations in both modules are presented as-is, with grading indicators. There is however room for extending this grading system so that the different levels become clearer. A simple solution to the credbility issue is to link grade letters and level indicators to additional information as found at Helsebiblioteket.no (Helsebiblioteket, 2013). In GAS, this information could reside on the guideline page, below the actual recommendations. Links from the grading levels could then lead the users down to the relevant background information. Participants relied on additional experience and knowledge to make decisions, and used the DIPS CDSS as a supplement. Few participants did however describe the system as untrustworthy, and their prior experience often included the same guidelines as used in this experiment. These guidelines, amongst others, are curriculum in medical studies at NTNU.
It is generally important to utilize critical thinking when using CDSSs and other support tools for decision making. This results in clinicians making their own decisions, and these decisions may differ from suggestions by a CDSS. A well-designed CDSS gives clinicians the option to choose, reflect and evaluate decisions suggested by the CDSS. Participants also mentioned that they need to use the AGR module over a period of time in order to fully trust the ranking. Only then are participants able to see a pattern in the ranking system, and therefore get some understanding of underlying ranking algorithms.

Chapter 11

Limitations

As with any software development project, this one also hit some obstacles during the development process. This chapter first look into possible improvements of the development process. It then dives into the experiment design, and point out different elements which could have been done differently.

11.1 Prototypes

This chapter lists those limitations related to implementation choices in both modules, as well as limitations of the DIPS system. During the implementation stage, several key choices were made. These choices greatly impacted on the module usability in both modules, and selected solutions as well as discarded solutions are presented here.

11.1.1 API Access vs Scraping

Guidelines from the Norwegian Electronic Health Library are essential for AGR and GAS, and obtainment of them were therefore the first and most important task to complete. Representatives of Helsebiblioteket.no were contacted in order to receive access to some API for fetching guidelines. While waiting for access, a script were made (see chapter 7). The purpose of this script is to scrape and parse guidelines, as well as to add metadata for easier access. Helsebiblioteket.no came through and sent source files containing guidelines, however they did not provide any API. The received files also contains less metadata than the generated files, and the official source files were scrapped.
If this had been a larger project than just a basis for an evaluation of a theroretical system, then the first objective would be to create a common API. This enables other projects to access and utilize the guidelines without the hassle of scraping

and modifying them while also adding identifying and meaningful metadata.

11.1.2 Internet Explorer 6 vs Chrome

Both AGR and GAS were originally developed to work with the Chrome web browser and the Chrome browser plugin. The new DIPS system uses the Chrome plugin, but lacks other features critical to the usability experiment. Therefore, it was necessary to use the current (and old) DIPS system for the tests. One advantage of using the current system, is that test patients and other data are pre-stored here due to earlier tests done in the EviCare project. A great disadvantage is that this DIPS system uses a fairly outdated and old browser plugin; Internet Explorer 6 (IE6). IE6 is subject to much criticism, and even Microsoft drives to stop users from using this version. IE6's market share has been estimated to 6,7% per april 2013 (Microsoft, 2013), and keeps falling.

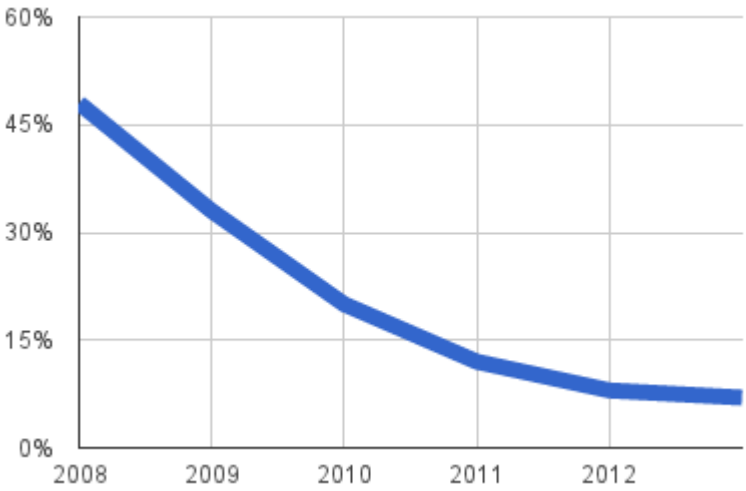


Figure 11.1: IE6 market share from jan. 2008 to jan. 2013 (source: NetMarket-Share (2013))

IE6 have little to no support for several web technologies like JavaScript or HTML5. This disabled most dynamic features in AGR, and some features provided by Google Custom Search in GAS. The original AGR module had features like expanding and collapsing guidelines, and functions to redefine guideline order for the test and styling. The most prominent drawback of poor styling (CSS) support were shown when subchapter headings failed to indent. That, amongst many other more subtle styling flaws may have made the module less appealing and usable.
Of course, there were multiple solutions to this compatibility problem. Three of those were considered realizable in this project;

- Install a plugin called Chrome Frame simulating Chromium in Internet Explorer (IE)
- Upgrade the IE plugin to a modern version
- Rewrite the code to comply with older versions of IE

11.1.2.1 Chrome Frame

Chrome Frame were made after years of adapting certain web sites to outdated versions of IE, primarily found in larger organizations in governments or private sectors. It is basically a plugin for old IE browsers, which emulates a Chrome browser inside IE. This means that developers only have to test their web applications on a limited number of semi-cross-compatible browsers, in theory. Or as Google describes it themselves:

> *Google Chrome Frame is an open source plug-in that seamlessly brings Google Chrome's open web technologies and speedy JavaScript engine to Internet Explorer.* (Google, 2013)

The HTML code were altered to support Chrome Frame, and to give users the ability to install Chrome Frame. This did however not work, even after some debugging.

11.1.2.2 Upgrade Internet Explorer plugin

DIPS were contacted in order to see if an upgrade to the IE plugin in DIPS were possible. After some consideration, this solution were dropped. An upgrade would probably consume time from an already tight schedule. Bugs could emerge by trying to integrate a modern version of IE in an outdated version of DIPS. This must however happen sooner or later if this DIPS system is to be used in future projects.

11.1.2.3 Code Refactoring - The Chosen Solution

This was deemed the most realizable solution, since no external effort were necessary. At first, the plan was to rewrite the JavaScript code to fit IE6's specifications. On the other hand, the list of incompatible JavaScript and JQuery elements were long, and embraced most features. AGR did not work at all, while GAS worked to some degree. The easiest solution were to exclude all JavaScript and JQuery code in AGR, and write a temporary static version with simple HTML.
The plan were successful, the module did however become very static without

several features that could have increased overall usability. As mentioned in chapter 9, participants wished for features that could have been implemented more effortlessly using a newer browser. That is, features like expanding and collapsing guidelines, as well as a dynamic menu or mini map of the site.
Not only AGR suffered from IE6 compatibility problems, GAS had some related bugs as well. The most prominent one were that users are unable to use the enter key after entering search queries into the search box. Results were also erased when the user alternated between guideline info pages and the result page.
Both AGR and GAS suffered from "black outs" as well, where the whole window for external patient information went blank and had to be restarted. That significally affected the participants' workflow.

11.1.2.4 Testing

Test routines outside the actual experiment are subject to criticism. One of two things could have been done from the project start. Either test the modules in IE6, or upgrade the IE6 plugin to IE10. The latter is as mentioned probably more time consuming than the former, and DIPS developers are understandably focused on creating the new refurbished version. Proper test routines could have cut down the development time significantly, while improved and increased communication with DIPS could have cut it down even further.

11.2 Experiment

The actual experiment also had some limitations and potential for improvement. The design featured some trade-offs between facilitating new users and facilitating more experienced users.
Participants were as mentioned in chapter 8.7 introduced to the think aloud method and made aware of the experiment goals. This prepping should have been done more thoroughly, as some students forgot to think aloud, and a minority also seemed to miss what the goal was. One participant had to cancel task 1 in the usability test prematurely, due to struggles related to low experience with the domain and clinical practice in general. That participant had to constantly be reassured of the tests purpose; to test the system, and not the user. This is a key principle in usability testing, to calm the participants down when they experience difficulties.

11.2.1 Tasks

Task information were also conveyed textually, and this information should have been more thorough and detailed. More explicit information about think aloud and experiment goals could also have been added to the set of tasks, in addition to the oral and textual information given before task execution. This would act as a reminder, making the information easier to process.
One task also made participants search and find the window for external patient information, which proved to be a difficult task. This task could have been excluded, as it had no part in the final evaluation. On another note, it introduced the DIPS system to participants, making it possible for them to familiarize with the user interface. Terje Røsand (2012, chapter 4.2) goes more in-depth in the description and analysis of these DIPS usability limitations.

11.2.2 GAS and AGR Order

All participants used GAS first (in task 1), before switching to AGR in task 2. This order could have made the AGR system seem more usable, as the participants get more acquainted with the surrounding user interface. Nevertheless, participants scored AGR much lower in the system usability scale forms. In order to avoid such uncertainties as this one, the experiment could have contained a much larger scaled usability test. This new version of the usability test should have contained a default system for guideline obtainment, without using either module. This default module would act as a basis for evaluating both AGR and GAS.
Another possibility is to add more participants and collect more data, while also having the opportunity to randomize tasks among them. Randomization is often used in clinical trials, and in this experiment, it would reduce aforementioned issues created by participants recalling steps from previous tasks. (Sauro, 2004) In practice, this would mean that participants use AGR, GAS or default DIPS in random order. More participants is however more time demanding, since more time have to be used to perform the experiment, in addition to recruitment and analysis.
AGR may also incorporate randomization, by randomizing between a set of correctly ranked guidelines and a set of ranked guidelines where the topmost guideline is less relevant than the following guidelines.

11.2.3 Patient Case

The patient case were rewritten and based on earlier work in the EviCare project. Some words were changed in order to accommodate inexperienced participants, the participant group consisted mostly of students with little experience. Cholesterol

lowering treatment were mentioned in the patient case text, which may have guided some participants more than necessary.

11.2.4 System Usability Scale

When answering the SUS forms, the participants were told to consider the whole DIPS system, they should instead have considered the actual modules, and tried to separate them from other DIPS elements. This would have gained more useful data, and could have been easy to integrate with a solution with more participants and randomization. SUS issues extend into interviews, since participants have their SUS answers fresh in mind when performing the interviews. When answering interview questions regarding system usability, some participants had a hard time focusing on just the modules, when they had to focus on the whole DIPS system in the previous SUS forms.

11.2.5 Participants

Timing around participant recruitment could have been planned better, meaning that students should have been provided a form of schedule where they could have marked their availability. Doodle (2013) and similar services provides exactly this feature, with little effort. As it was done now, the students enrolled first via e-mail, without specifying their available slots in their schedule, and then specified that in a later e-mail. This back and forth made it necessary to postpone the usability test with one working day. However, this had only minor additional effects on the experiment as a whole, only potentially losing some of the more busy students (mostly higher grade students).

Chapter 12

Conclusion

This chapter completes the discussion chapter by presenting the research questions, providing answers where possible, and evaluates the process of obtaining these answers.

12.1 Research Questions

The research questions presented in chapter 4.3 provides a basis for summarization of results discussed in chapter 10: *Discussion* and chapter 11: *Limitations*.

RQ1 How can a CDSS best utilize patient information as a context for ranking relevant recommendations?

This project shows two different ways of navigating CDSSs; search-based navigation and content-based automatic access to guidelines. The first, dubbed **Guideline Access using Search** gives users a simple interface for accessing guidelines without any additional means of navigation. The other, dubbed **Automatic Guideline Ranking** took results from a ranking algorithm and displayed the results with little user interaction. These two modules each represent an extreme point in navigational spectrum shown in figure 12.1.

Figure 12.1: Navigational spectrum, with GAS and AGR as two extremes

Several other solutions may emerge from these two, whereas one possible outcome is a combination of both modules. This combination could show guidelines automatically based on patient information from an EHR system, while giving the user the opportunity to search for other relevant and non-relevant guidelines. A new hypthesis is that this combination will gain better user satisfaction results. Increased user satisfaction increases adoption of CDSS by clinicians, and computerized decision making offers guidelines at the point of care, which in turn may increase quality of patient care.

RQ1.1 Which of search-based or content-based recommendation ranking gives best user satisfaction?

It seems like participants in this test struggled to some degree to find the guidelines they were looking for in the search-based solution, and prefer solutions which automate guideline access. They seem to prefer the automated solutions when the module is considered as a stand-alone system. The automated solution greatly improves availability of guidelines from the search-based solution, which in turn creates a higher level of awareness for clinician users.
Clinicians prefer a solution with actions requiring little back and forth between different states or windows. External CDSS solutions outside the EHR system does require a lot of alteration between different systems with different designs, which decreases consistency and in turn user satisfaction. Integrating CDSS with EHR systems enables a smoother transition between patient information review and decision making.
These result may of course be applicable only to inexperienced medical students, and further research must be performed in order to explore how experienced clinicians utilize search and content based access to clinical guidelines compared to inexperienced clinicians.

RQ2 How does CDSS integrated in the EHR system affect clinical workflow?

This question is best divided and answered in two sub-question, where one sheds light on user repsonses to this new tool, and another which compares this computerized integration to manual solutions. Manual solutions and solutions with external guideline collections dominates clinical workflows today, as mentioned in chapter 5.

RQ2.1 How does clinical users respond to integrated decision support in the EHR system?

This project focused on integrating the CDSS prototypes with an existing EHR system. Responses from interviews show that clinical users prefer integration between CDSS and EHR systems, over solutions which require the use of several conflicting and inconsistent external sources.
The EHR system were not tested without any CDSS integration, nor were any external CDSS tested. Scores from the SUS forms show that GAS integrated with DIPS EHR system gains a higher score than DIPS integrated with AGR (average SUS score of 42 with GAS and 24 with AGR). These SUS results does not however conclude that either one of these combination gives better user satisfaction, as several dependent factors may affect the results. Factors include DIPS usability and the fact that participants learn how to use this EHR system as they use it.
The most reliable data here is data from the usability testing and reflection in the interviews after each prototype test. Data here clearly suggest that participants prefer automatic access to guidelines in a context of patient information, but this depends on factors such as credibility and awareness. Credibility meaning that they trust the recommendations given to them, and trust that these recommendations are indeed relevant to the selected electronic health record. Awareness means that an integrated CDSS displays clinical recommendations at the place and time of decision support. An integrated solution clearly displays recommendations at the *place* of decision support, and timing is decided by when the recommendations become available for clinicians in the integrated system. Both prototypes in this scenario gave recommendations at the time of decision making, but results show that search-based solutions may be more cumbersome and therefore more time consuming.
These results are not generalizable, due to a small and somewhat homogenous participant group. Experienced clinicians may (and probably will) give differing answers to some degree. Chapter 13 discusses how future research and prototype implementation may uncover more details around this research question.

RQ2.2 How does automated computerized decision support compare to manual decision support?

Participants in this experiment stated that they clearly prefer computerized decision support over manual decision support. These participants were fairly young (all in their 20's), and rather inexperienced. Preliminary survey suggest that they almost always access guidelines through computerized CDSS, and their age suggest that most of them are fairly used to digital information retrieval.

12.2 Process Evaluation

I originally had no prior experience with health informatics, so this master project started with a steep learning curve. The process of studying this domain helped me learn a lot about health informatics in general and more specifically clinical practice, decision support, medical information retrieval, electronic health records and much much more.
The process also included research of methods such as experiments and case-studies, where I got to use what I've learned about usability and usability testing throughout my bachelors and masters degree. My supervisor and Terje Rødsand (amongst others in EviCare) helped design and execute the case-based experiment including recruitment of particiants and usability testing at the usability lab. Their help made managing the experiment considerably easier.

Chapter 13

Future Work

The modules are merely prototypes, and need additions and changes in future iterations before finalization. This chapter presents different solutions to the previously mentioned limitations and other changes based on participant feedback. In addition to interface and implementation alterations, it would be benefitial to adjust the actual test experiment. The first subchapters presents suggestions for future module improvements, while the following subchapter presents an alternative module, which comprise of features from both modules. To round up this chapter, suggestions for future experiments are presented.

13.1 Guideline Access using Search

Future work with the GAS module include improvements to the search interface inspired by other similar systems, and inspired by participant feedback. Some of these improvements are both common in other similar websites, while also being suggested by participants. The reason for this interception is mostly because the participants are familiar with several websites which look and feel like the GAS interface.

13.1.1 Advanced Search

In order to give the clinicians more control over the results, one participant suggested to add advanced search. In this case, it means to add scoped search. Scoped search enables the users to narrow down the search scope by defining a medical domain. If a future version of GAS were to include several other medical domains in addition to the stroke guidelines, it would benefit the clinicians if they could select the domain by themselves. The medical search engine UpToDate (2013) employs such a feature, as mentioned earlier. Future versions of GAS may add

domain selection from the front page, and the default domain will be all guidelines available from the Norwegian Medical Health Library. Advanced search may also contain other features like grade selection. But additional features involves more choices for the clinicians, and should only be available after they have done the initial simple search. Nielsen (2001) points out that simple search is the preferred method, as studies show that users search with a mean query length of two words. The same article states that users most likely will give up search after the first, second or third try. Most users gave up after the first try. This shows the importance of creating low barriers, while also aiding the users after their initial search. Based on this research, one could also see how important it is for user satisfaction to produce relevant results from the first search, and the article says it best:

> Another reason to emphasize early success is that users typically make very quick judgments about a website's value based on the quality of one or two sets of search results. If the list looks like junk, they may abandon the site completely. At a minimum, they'll forgo the site's search in favor of external search engines like Google. (Nielsen, 2001)

Another possibility is to suggest other searches for clinicians after their initial search, similar to what Google Search does. Googles version is often used for spell checking (Google, 2013), but also find synonyms to frequent terms. When clinicians fail to remember the exact phrasing of a guideline, they may search for semi relevant guidelines or terms that are either directly or indirectly related to terms in the wanted guideline. Then the GAS module could produce a suggestion for another guideline related to what the clinicians searched for, as shown in figure 13.1.

13.1.2 Wiki Inspiration

In addition to the advanced search, participants suggested another solution which draws inspiration from the wikis of the web, like Wikipedia. A wiki is a website created and maintained through a community of people, with pages of different topics linked to each other. Unlike wikis, CDSS information is not provided by communities, but by professional clinicians. The suggestion is about taking the topic-linking and adding that to guidelines in GAS, namely links to guidelines relevant to the selected one. But links to other external information, like the suggested grading background information, may also be provided.

13.1.3 Frequently Searched Terms

GAS does not work optimally when clinicians have to restart their search after reviewing a guideline. The search now "deletes" the results list, which stems from

Figure 13.1: GAS with search suggestion ("Did you mean?")

IE6 compatability problems (see chapter 11.1.2). This bug creates a situation where clinicians have to keep track of their searches at all times, and future versions will of course remove this issue. Another way of increasing user satisfaction in this case, is to add a list of frequently searched terms to the front page. These terms could either be represented as links (which, when clicked, performs automatic search) or just by suggestions when clincians starts to type. The former lowers the barrier for clinicians while taking away some of their control, but also creates another element on the front page. The latter solves the problem of front page clutter, but will have to provide relevant suggestions to increase usability. The preferred solution is automatic search, as long as it gives clinicians useful suggestions. Both solutions should enable users to clear search history, as some of the searches includes sensitive information. An option to clear history also increase system trustworthyness.

Figure 13.2: GAS with frequently searched terms

13.1.4 Other Suggestions

Other notable suggestions also emerged, as seen in the Results chapter. One participant suggested to drop the paged result list and design a solution with a single page design. This involves a more dynamic approach to view results. Results

are loaded as the user scrolls down the result page, and does not require users to click links for the next page of results. One Page Love (2013) lists several sites which incorporates this idea.
Google Custom Search is not especially designed for searching through clinical guidelines, but enables customization. The goal for a potential new underlying search engine would be to train or customize it so that it adapts to the clinical workflow. This means better support for acronyms common in medicine, as well as synonyms and other features like spelling suggestions for when clinicians spell common clinical terms wrong. This helps holding on to clinicians as users, and ensures that GAS and the DIPS CDSS avoids bad reputation. A bad reputation could destroy further attempts of improving the module, as clincians most likely would fall back to and prefer their usual rythm and tools that function as expected in their current workflow.

13.2 Automatic Guideline Ranking

AGR features a very simple and static design, and participants had few suggestions for additions or alterations to the design. The original AGR design before IE6 optimalizations are more dynamic and easier to look at with more styling. DIPS will be replaced by a simpler and more elegant system, which incorporates a Chrome plugin for modules like AGR and GAS. This plugin does support several of the styling properties from the original design, as well as other dynamic features like expanding and collapsing guidelines. Guidelines are then presented just as headings, which reveal selected guidelines as the user clicks those headings, by expanding the guideline table. Another missing design element is seperation of elements by indentation and coloring. Guidelines are categorized, and the individual guidelines are then indented to indicate which parent category it belongs to. Coloring indicates what is clickable and what is not, as almost all other websites implement it.

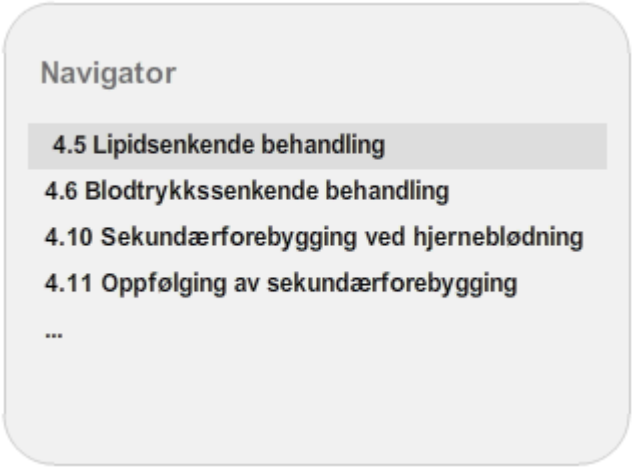


Figure 13.3: Mockup of a navigational menu for AGR

The most common suggestion is to add some sort of centralized list or menu of clickable guideline headlines, as shown in figure 13.3. This list will be placed at the top, besides the topmost guideline. By placing it there, it captures clinicians attention. This prevents **lack of awareness**, and therefore solves a significant issue with the current design.

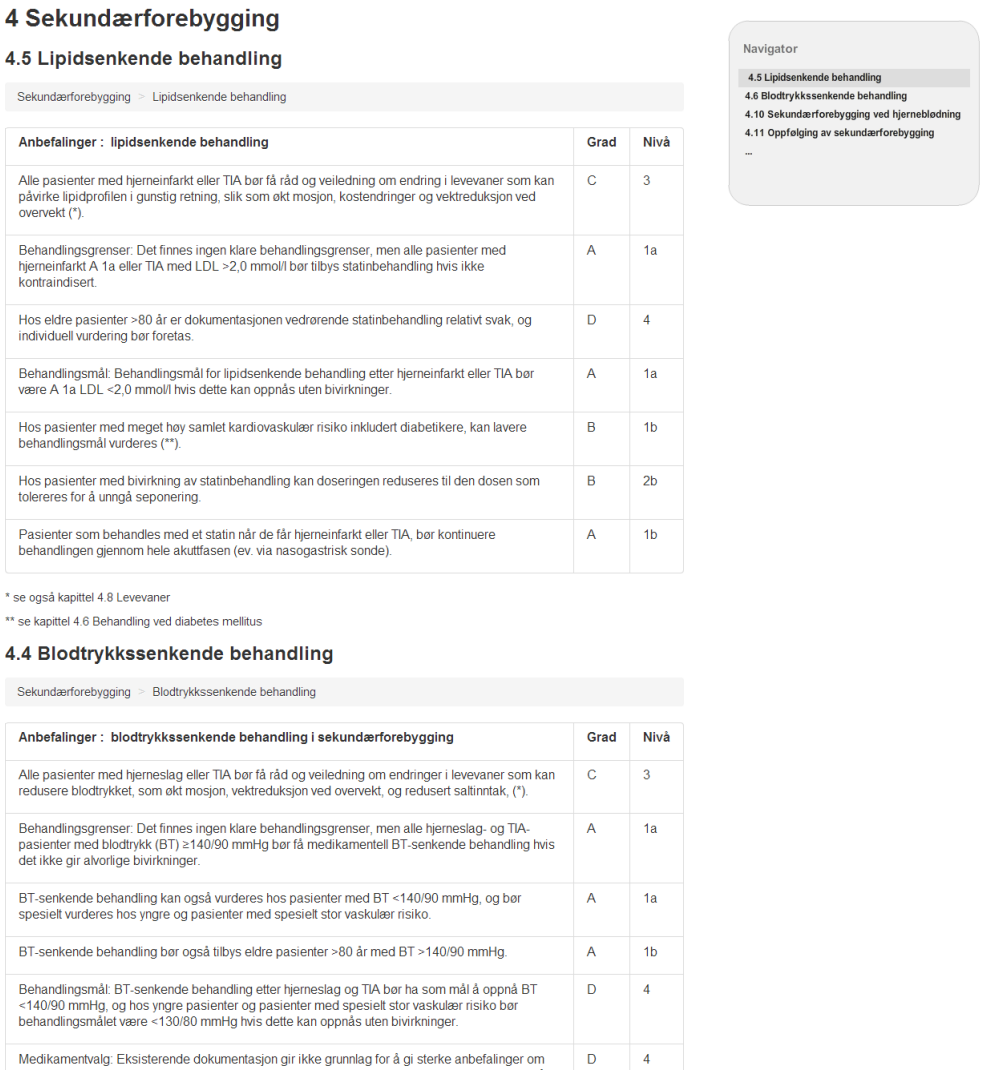


Figure 13.4: Mockup of AGR with navigational menu

13.3 Combining AGR and GAS

Participants also mentioned during the interview that although they preferred AGR, a combination of both AGR and GAS would suit their needs best. Some of them may have used solutions which combine both browsing and search, as many websites take advantage of both solutions. Already when using the GAS module, some participants requested a way of browsing the different categories in the stroke guidelines. They noticed that guideline headings featured a identifying number similar to chapter numbering in books (eg. 4.5, 4.6 and so on), and

wanted to browse parent categories to search for similar guidelines to the selected one. These numbers stem from Helsebiblioteket.no, which some of the participants may be familiar with (almost all participants answered yes when asked if they had used clinical guidelines before).
In addition to requesting simple browsing features in GAS, many participants wished for search functionality in AGR. They had at that point used and gotten familiar with GAS, which may have affected their feedback (see chapter 11.2.2). One participant did however try to open a search interface by using the well known shortcut Ctrl+F, which often opens a search box to search through the current text in multiple browsers and text editors. As mentioned in 10.1 *Guideline Access using Search*, usability usually increase when users recognize website elements from similar websites. It is natural to compare these modules to Helsebiblioteket.no, which features a tree structure and search functionality. This website is however not alone in implementing search as browsing, search is taking over the web (see chapter 5).
A combination of AGR and GAS would preferrably show relevant guidelines automatically, and then give the option to search through both through the domain of the listed guidelines, as well as other domains. The final module will act much like Helsebiblioteket.no, but with a simpler design and interface for integration in CDSSs like DIPS. This module would necessary also have to address credbility issues mentioned in chapter 10.3. Some clinicians may be skeptical, but by implementing support for better research background information, some of these issues will go away. Solutions range from adding useful and relevant information on the research work behind each guideline adjacent to the selected guideline, or as links from the guideline (with an easy way of going back).

13.4 Further EHR Integration

The next step after combining both modules into one "supermodule", could be to integrate the module further into DIPS. This will of course make it less portable, and make integration into competing CDSSs more cumbersome. Further integration would mean to open the module at relevant point in time during clinical workflows, without distracting the clinicians. Subtle links from the EPJ could lower the barrier between patient journals and clinical guidelines. Tagging of metadata in these journals is the optimal way of making that work, preferrably with automatic tagging. Several subprojects in the EviCare (2013) project involves just that, and future iterations of the guideline browsing interface will take advantage of their outcomes.

13.5 Future Experiments

After redesigning the modules, the next step is redesigning the actual experiment design. A future experiment should include a wider target audience, in addition to increasing the participant group. Only then can the experiment produce results based on the experience and knowledge level of participating clinicians. Lack of agreement and intertia of previous practice should be measured to a greater extent. Comparing feedback from inexperienced participants (namely medical students) with experienced clinicians may produce interesting results. The hypothesis here is that inexperienced students prefer less user interaction and more automation than experienced clinicians. They also seem to trust the decision support system more than their experienced counterparts. Results from this experiment suggest that there is a difference, but these results are not significant enough due to the participant distribution (over half of the group consist of students from the third year, see figure 9.1).

The automatic ranking system heavily depends on the development progress of either AGR or the combined module. The current experiment did not include actual automatic ranking, but featured a pre-ranked list. The list did however not rank all guidelines as expected by the participants. Future versions of the experiment should re-rank the list if it is to be used instead of an actual ranking system.

Participants stated that they would need to use the AGR module multiple times in order for them to trust the results. A more usable module would most likely remove this issue, as the learnability also increases. Future experiments may incorporate some way of repeated testing, where participants use the module(s) in medical practice. This experiment design is also dependant on how the next module iteration works. "Cheating" by using pre-ranked lists is unfeasible in an experiment where observators have little or no control over the environment. One could also combine both experiments by setting up a close-to-real sitation in the usability lab at NSEP by using artifacts like hospital equipment, beds and extras.

Bibliography

Alpert, J. and Hajaj, Nissan, G. W. I. T. (2008). We knew the web was big... `http://googleblog.blogspot.no/2008/07/we-knew-web-was-big.html`. Accessed: 2013-05-12.

Baeza-Yates, R. and Ribeiro-Neto, B. (2011). *Modern Information Retrieval*. Addison-Wesley, second edition.

Berlin, A., Sorani, M., and Sim, I. (2006). A taxonomic description of computer-based clinical decision support systems. *Journal of biomedical informatics*, 39(6):656–67.

Boxwala, A. A., Tu, S., Peleg, M., Zeng, Q., Ogunyemi, O., Greenes, R. A., Shortliffe, E. H., and Patel, V. L. (2001). Toward a representation format for sharable clinical guidelines. *Journal of biomedical informatics*, 34(3):157–169.

Brooke, J. (1996). Sus: A quick and dirty usability scale, pages 189–194. usability evaluation in industry.

Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P.-A. C., and Rubin, H. R. (1999). Why don't physicians follow clinical practice guidelines? *JAMA: the journal of the American Medical Association*, 282(15):1458–1465.

Cutting, D. R., Karger, D. R., Pedersen, J. O., and Tukey, J. W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 318–329. ACM.

Doney, L., Barlow, H., and West, J. (2005). Use of libraries and electronic information resources by primary care staff: outcomes from a survey. *Health Information and Libraries Journal*.

Doodle (2013). Doodle: easy scheduling. `www.doodle.com`. Accessed: 2013-05-12.

EviCare (2009). Evidence-based care processes: Integrating knowledge in clinical information systems. page 13.

EviCare (2013). EviCare – Evidence-based care processes: Integrating knowledge in clinical information systems. http://www.kunnskapssenteret.no/Prosjekter/EviCare+%E2%80%93+Evidence-based+care+processes%3A+Integrating+knowledge+in+clinical+information+systems.8561.cms. Accessed: 2013-05-16.

for Chronic Conditions (UK), N. C. C. (2008). National clinical guideline for diagnosis and initial management of acute stroke and transient ischaemic attack (tia).

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P., Beyene, J., Sam, J., and Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA: the journal of the American Medical Association*, 293(10):1223–1238.

Google (2013). Google Chrome Frame. https://developers.google.com/chrome/chrome-frame/. Accessed: 2013-05-12.

Google (2013). Google Search Features. https://www.google.com/help/features.html#keyword. Accessed: 2013-05-23.

Google Custom Search (2013). Google Custom Search. https://www.google.com/cse/all. Accessed: 2013-05-12.

Gunter, T. D. and Terry, N. P. (2005). The emergence of national electronic health record architectures in the united states and australia: models, costs, and questions. *Journal of Medical Internet Research*, 7(1).

Hayward, R. (2004). Clinical decision support tools: Do they support clinicians? *The Canadian Medical Association Journal*, 170(10):66–68.

Healthcare Improvement Scotland (2013). Scottish Intercollegiate Guidelines Group. http://www.sign.ac.uk/. Accessed: 2013-05-23.

Helsebiblioteket (2013). Gradering av kunnskap og anbefalinger. http://www.helsebiblioteket.no/retningslinjer/hjerneslag/metode-og-kunnskapsgrunnlag/gradering. Accessed: 2013-05-14.

Hill, K. (2008). Australian clinical guidelines for acute stroke management 2007. *International Journal of Stroke*, 3(2):120–129.

Hunt, D. L., Haynes, R. B., Hanna, S. E., and Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *JAMA: the journal of the American Medical Association*, 280(15):1339–1346.

Iso (1998). Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability.

Kawamoto, K., Houlihan, C. a., Balas, E. A., and Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ (Clinical research ed.)*, 330(7494):765.

Lappa, E. (2005). Undertaking an information-needs analysis of the emergency-care physician to inform the role of the clinical librarian: a greek perspective. *Health Information and Libraries Journal*, 22(2):124–132.

Lewis, C. (1982). *Using the" thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center.

Lewis, C. and Rieman, J. (1993). Task-centered user interface design. *A Practical Introductio*.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140:1–55.

Microsoft (2013). Internet Explorer 6 Countdown. `http://www.ie6countdown.com/`. Accessed: 2013-05-12.

NetMarketShare (2013). Market Share Statistics for Internet Technologies. `http://www.netmarketshare.com/`. Accessed: 2013-05-23.

Nielsen, J. (2001). Search: Visible and Simple. `http://www.nngroup.com/articles/search-visible-and-simple/`. Accessed: 2013-05-14.

Nielsen, J. (2011). Top 10 Mistakes in Web Design. `http://www.nngroup.com/articles/top-10-mistakes-web-design/`. Accessed: 2013-05-14.

Nielsen, J. (2012). Usability 101: Introduction. `http://www.nngroup.com/articles/usability-101-introduction-to-usability/`. Accessed: 2013-05-12.

Oates, B. J. (2006). *Researching Information Systems and Computing*. SAGE Publications Ltd.

One Page Love (2013). One Page Love. http://onepagelove.com/. Accessed: 2013-05-16.

Placebo (2013). Mf placebo. http://org.ntnu.no/placebo/. Accessed: 2013-05-12.

Python (2013). Python Programming Language. www.python.org/about. Accessed: 2013-05-12.

Sauro, J. (2004). The Importance Of Task Order Randomizing During A Usability Test. http://www.measuringusability.com/random.htm. Accessed: 2013-05-12.

Scott, H. and Fairweather (2000). The use of evidence-based medicine in the practice of consultant physicians. results of a questionnaire survey. *Australian and New Zealand Journal of Medicine*, 30(3):319–326.

Shariff, Bejaimal, S. I. W. H. S. T. and Garg (2011). Searching for medical information online: a survey of canadian nephrologists. *Journal of Nephrology*.

Terje Røsand (2012). Think aloud methods with eye tracking in usability testing: A comparison study with different task types. Master's thesis.

The Norwegian Knowledge Centre for the Health Services (2004). The Norwegian Knowledge Centre for the Health Services. http://www.kunnskapssenteret.no/Home. Accessed: 2013-05-12.

The Norwegian Knowledge Centre for the Health Services (NOKC) (2013). Retningslinjer - Helsebiblioteket.no. http://www.helsebiblioteket.no/retningslinjer. Accessed: 2013-05-12.

Tobii (2013). Eyetracking for research. http://www.tobii.com/en/eye-tracking-research/global/. Accessed: 2013-05-28.

Twitter (2013). Bootstrap. http://twitter.github.io/bootstrap/. Accessed: 2013-05-23.

UpToDate (2013). UpToDate.com. http://www.uptodate.com/contents/search. Accessed: 2013-05-12.

US National Library of Medicine (2013). PubMed - National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/pubmed/. Accessed: 2013-05-12.

World Wide Web Consortium (1999). XML Path Language. http://www.w3.org/TR/xpath/. Accessed: 2013-05-23.

Yahoo (2013a). Yahoo Directory. http://dir.yahoo.com/. Accessed: 2013-05-28.

Yahoo (2013b). Yahoo Query Language.

Zheng Wang (2012). Requirements elicitation on integrating clinical guidelines to electronic patient record: An empirical study. Master's thesis.

Appendix A

Forms

A.1 Tasks

The following is the full task text used in the usability tests.

Kliniske retningslinjer - brukbarhetstest

Om du ikke har fylt ut spørreskjema på forhånd, kan du trykke her. (Edit: This was a link to the form in chapter A.2)
Hensikten med denne brukbarhetstesten er å evaluere to metoder for å hente frem relevante kliniske retningslinjer. En retningslinje er et sett med utviklede faglige anbefalinger som etablerer en nasjonal standard for pasientbehandling. Det ble opprettet tre arbeidsgrupper med representanter oppnevnt av Helsedirektoratet. Arbeidsgruppene har utarbeidet det faglige grunnlaget for anbefalingene i retningslinjen etter internasjonal anerkjent metodikk for retningslinjearbeid.
På denne linken finner du mer informasjon om hvordan anbefalingene i hver retningslinje er gradert etter forskningsgrunnlag og dokumentasjon.

Oppgaver

Nå får du to sett med oppgaver. Du skal gjennomføre oppgavene kronologisk, og hvert sett med oppgaver har omtrent lik fremgangsmåte. Underveis mens du gjennomfører oppgavene, er det viktig at du verbaliserer tankene dine, altså at du tenker høyt. Både video og lyd vil bli tatt opp, samt øyebevegelser. Etter hvert sett med oppgaver, skal du besvare et spørreskjema, så gjennomfører vi et intervju mens går vi over videomaterialet.
Begge oppgavene omhandler samme (fiktive) pasientcase:

Pasienten er en mann født 20.06.1961, og fikk et hjerneslag for to år siden. På Rikshospitalet ble han behandlet for åpen foramen ovale. Han har vært på poliklinikken to ganger tidligere, og dette var det tredje besøket. Hans LDL kolesterolnivå var på 2,4.
I følge retningslinjene, skal alle pasienter som har fått hjerneslag og har et LDL-nivå over 2,0 bli tilbudt statinbehandling, noe som er et lipidsenkende legemiddel.

I første oppgave skal du teste en prototype på retningslinjesøk, hvor retningslinjene rangeres etter relevans til søket

Retningslinjesøk

1. Hent frem og velg pasient "Henriksen, Stein"
2. Forsikre deg om at den valgte pasienten nå er "Henriksen, Stein"
3. Finn epikriser på pasienten
    - (både epikriser og poliklinisk notat er oppført som epikriser i DIPS-systemet)
    - Du har ti minutter på å sette deg inn i epikrisene, før pasienten ankommer
4. Ved hjelp av ekstern pasientinformasjon skal du finne relevante retningslinjer
    - Åpne vinduet for ekstern pasientinformasjon
    - Søk deg frem til relevante retningslinjer basert på epikrisene du leste tidligere (bruk søk-knappen, "enter" fungerer ikke)

Fyll ut dette spørreskjemaet (Edit: Contained a link to the form in chapter A.3) om retningslinjesøk når du er ferdig med oppgaven
I oppgave 2 skal du teste en prototype med automatisert rangering og visning av retningslinjer basert på pasientinformasjon

Automatisert rangering

1. Hent frem og velg pasient "Havnut, Vidar"
2. Forsikre deg om at den valgte pasienten nå er "Havnut, Vidar"

3. Pasienten har samme sykdomsforløp som i forrige oppgave

    - Du har anledning til å lese epikrisene nok en gang

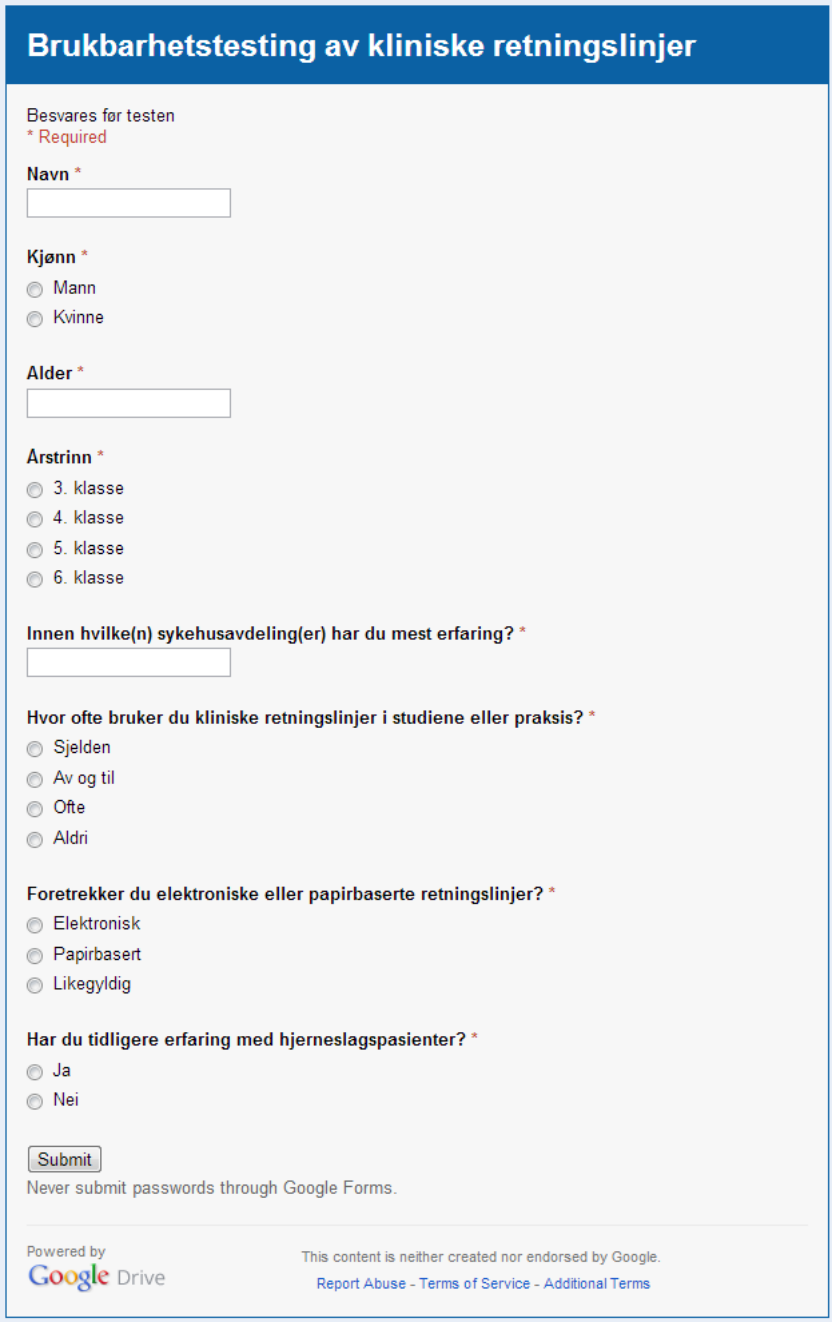
4. Ved hjelp av ekstern pasientinformasjon skal du finne relevante retningslinjer

    - Åpne vinduet for ekstern pasientinformasjon

    - Finn frem den mest relevante retningslinjen basert på epikrisene du leste tidligere

Fyll ut dette spørreskjemaet (Edit: Contained a link to another version of the form in chapter A.3) om automatisert rangering av retningslinjer når du er ferdig med oppgavene

A.2 Pre-study Form



Brukbarhetstesting av kliniske retningslinjer

Besvares før testen
* Required

Navn *

Kjønn *
- Mann
- Kvinne

Alder *

Årstrinn *
- 3. klasse
- 4. klasse
- 5. klasse
- 6. klasse

Innen hvilke(n) sykehusavdeling(er) har du mest erfaring? *

Hvor ofte bruker du kliniske retningslinjer i studiene eller praksis? *
- Sjelden
- Av og til
- Ofte
- Aldri

Foretrekker du elektroniske eller papirbaserte retningslinjer? *
- Elektronisk
- Papirbasert
- Likegyldig

Har du tidligere erfaring med hjerneslagspasienter? *
- Ja
- Nei

Submit

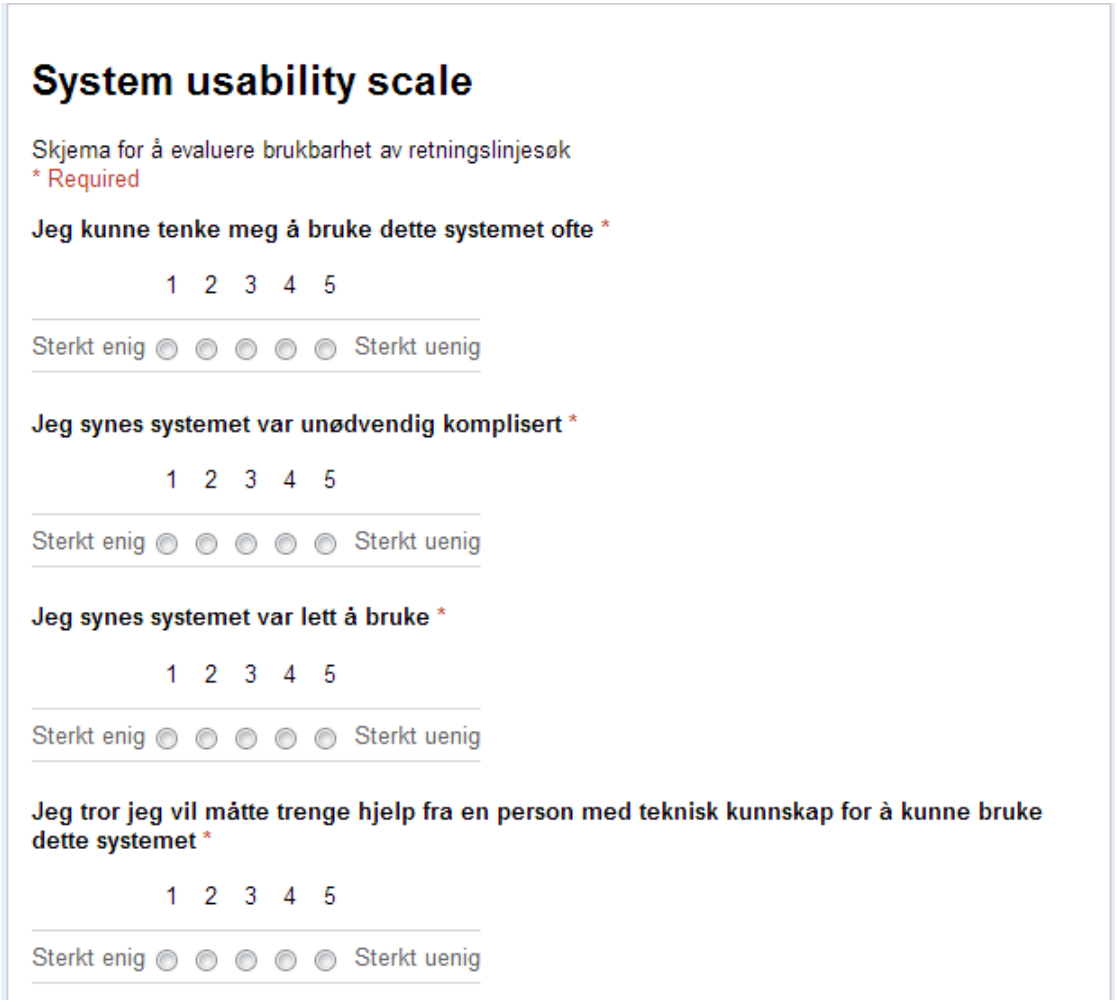
Never submit passwords through Google Forms.

Powered by Google Drive

This content is neither created nor endorsed by Google.
Report Abuse - Terms of Service - Additional Terms

Figure A.1: Form answered by participants before usability tests

A.3 System Usability Scale Form



System usability scale

Skjema for å evaluere brukbarhet av retningslinjesøk
* Required

Jeg kunne tenke meg å bruke dette systemet ofte *

1 2 3 4 5

Sterkt enig ○ ○ ○ ○ ○ Sterkt uenig

Jeg synes systemet var unødvendig komplisert *

1 2 3 4 5

Sterkt enig ○ ○ ○ ○ ○ Sterkt uenig

Jeg synes systemet var lett å bruke *

1 2 3 4 5

Sterkt enig ○ ○ ○ ○ ○ Sterkt uenig

Jeg tror jeg vil måtte trenge hjelp fra en person med teknisk kunnskap for å kunne bruke dette systemet *

1 2 3 4 5

Sterkt enig ○ ○ ○ ○ ○ Sterkt uenig

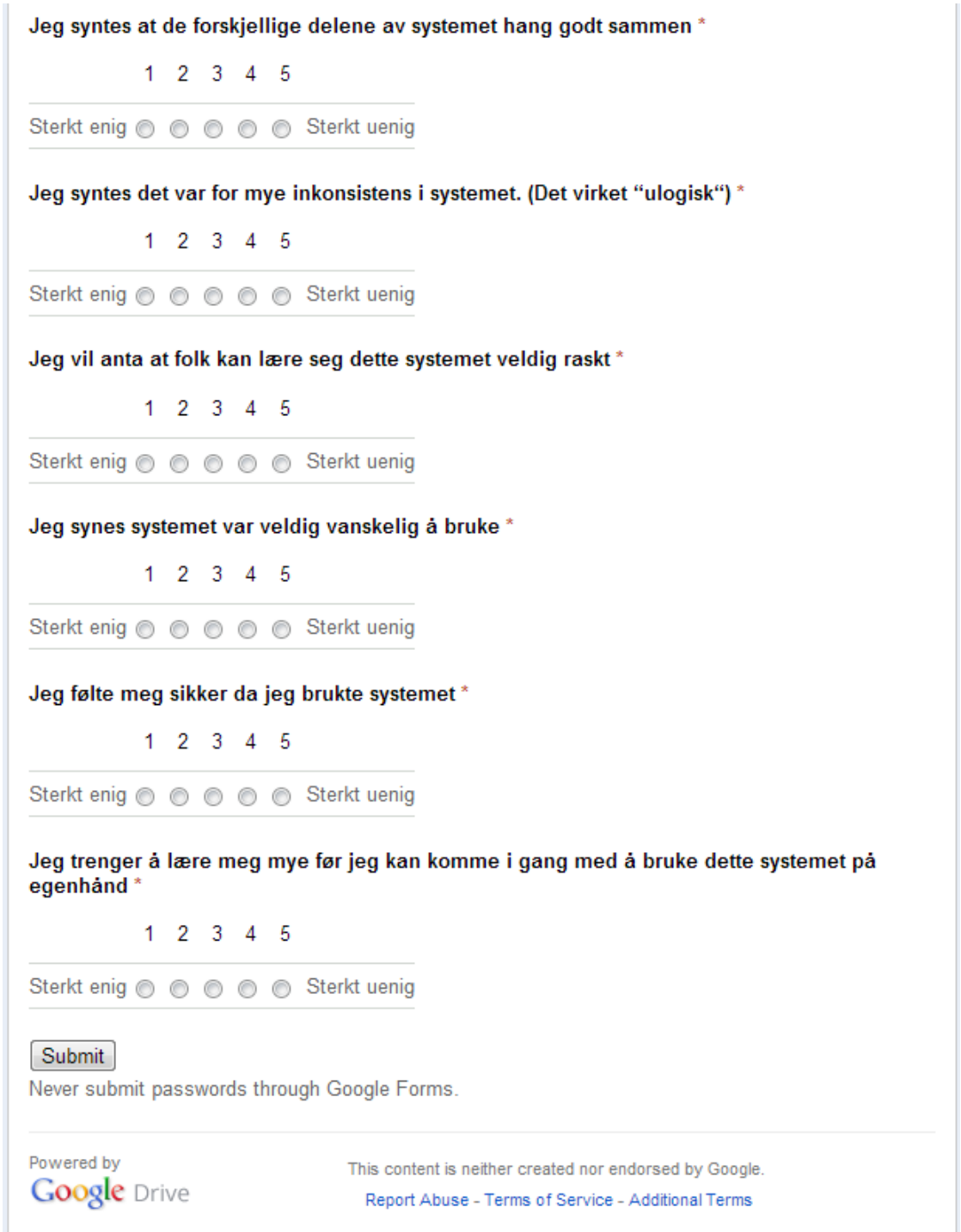


Figure A.2: System Usability Scale (SUS) form

Appendix B

Figures

B.1 Source code

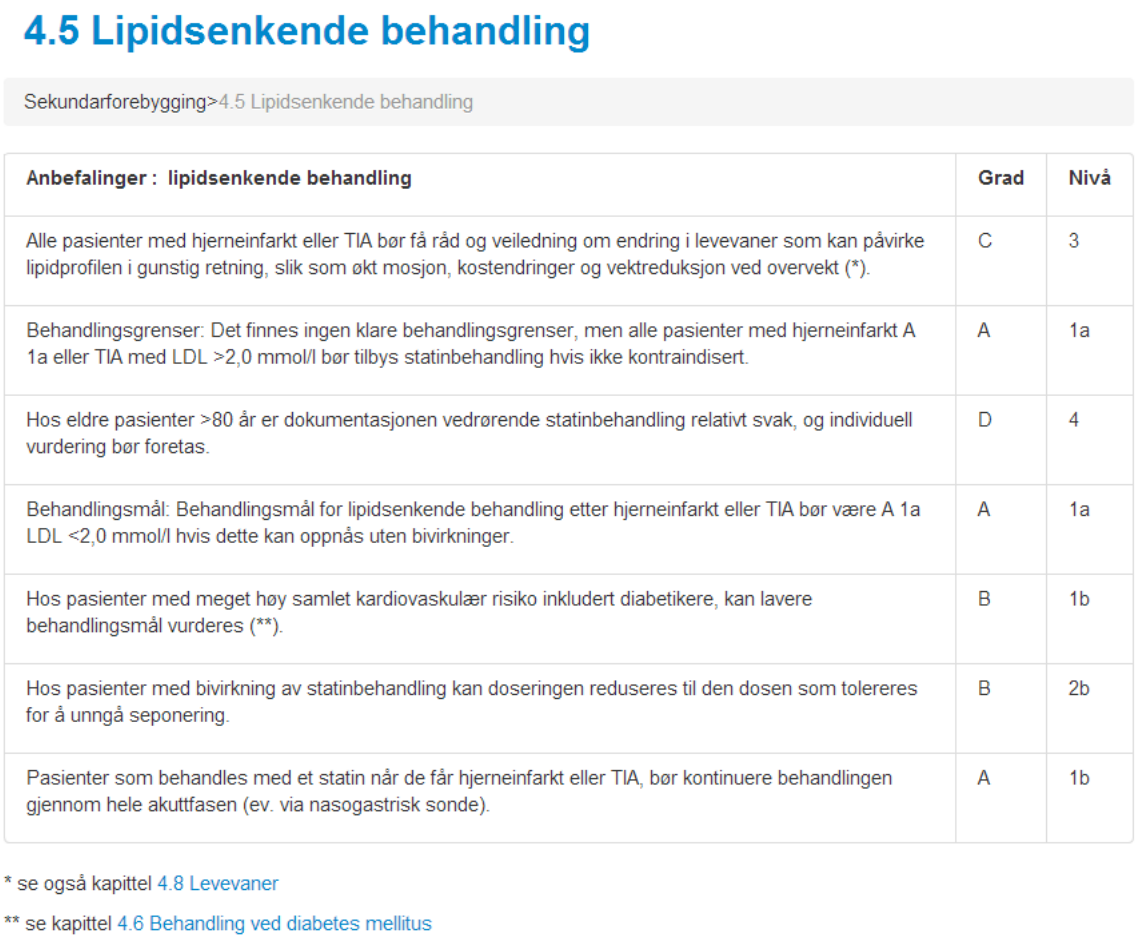
```php
<?php
    if ( strpos($_GET['patientId'], '1000270' ) !== FALSE ) {
        header ('Location: http://folk.ntnu.no/chriper/Retningslinjer/recommendations_ranked.html');
        exit;
    }
    if ( strpos($_GET['patientId'], '1000218' ) !== FALSE ) {
        header ('Location: http://folk.ntnu.no/chriper/Retningslinjer/search.html');
        exit;
    }
?>
```

Figure B.1: PHP redirect script

```xml
<guidelines>
  <recommendation id="recommendation0" link="urlgoeshere">
    <div class="breadcrumbs">
      <a href="urlgoeshere">Title</a>
    </div>
    <h2 class="chapterheading">Header</h2>
    <div id="recommendations">
      <table class="recommendation">
      </table>
    </div>
  </recommendation>
</guidelines>
```

Figure B.2: XML recommendation structure

B.2 The Prototypes



4.5 Lipidsenkende behandling

Sekundarforebygging>4.5 Lipidsenkende behandling

Anbefalinger : lipidsenkende behandling	Grad	Nivå
Alle pasienter med hjerneinfarkt eller TIA bør få råd og veiledning om endring i levevaner som kan påvirke lipidprofilen i gunstig retning, slik som økt mosjon, kostendringer og vektreduksjon ved overvekt (*).	C	3
Behandlingsgrenser: Det finnes ingen klare behandlingsgrenser, men alle pasienter med hjerneinfarkt A 1a eller TIA med LDL >2,0 mmol/l bør tilbys statinbehandling hvis ikke kontraindisert.	A	1a
Hos eldre pasienter >80 år er dokumentasjonen vedrørende statinbehandling relativt svak, og individuell vurdering bør foretas.	D	4
Behandlingsmål: Behandlingsmål for lipidsenkende behandling etter hjerneinfarkt eller TIA bør være A 1a LDL <2,0 mmol/l hvis dette kan oppnås uten bivirkninger.	A	1a
Hos pasienter med meget høy samlet kardiovaskulær risiko inkludert diabetikere, kan lavere behandlingsmål vurderes (**).	B	1b
Hos pasienter med bivirkning av statinbehandling kan doseringen reduseres til den dosen som tolereres for å unngå seponering.	B	2b
Pasienter som behandles med et statin når de får hjerneinfarkt eller TIA, bør kontinuere behandlingen gjennom hele akuttfasen (ev. via nasogastrisk sonde).	A	1b

* se også kapittel 4.8 Levevaner

** se kapittel 4.6 Behandling ved diabetes mellitus

Figure B.3: AGR module initial design with Bootstrap styling

4.5 Lipidsenkende behandling

- Sekundarforebygging>
- 4.5 Lipidsenkende behandling

Anbefalinger : lipidsenkende behandling	**Grad**	**Nivå**
Alle pasienter med hjerneinfarkt eller TIA bør få råd og veiledning om endring i levevaner som kan påvirke lipidprofilen i gunstig retning, slik som økt mosjon, kostendringer og vektreduksjon ved overvekt (*).	C	3
Behandlingsgrenser: Det finnes ingen klare behandlingsgrenser, men alle pasienter med hjerneinfarkt A 1a eller TIA med LDL >2,0 mmol/l bør tilbys statinbehandling hvis ikke kontraindisert.	A	1a
Hos eldre pasienter >80 år er dokumentasjonen vedrørende statinbehandling relativt svak, og individuell vurdering bør foretas.	D	4
Behandlingsmål: Behandlingsmål for lipidsenkende behandling etter hjerneinfarkt eller TIA bør være A 1a LDL <2,0 mmol/l hvis dette kan oppnås uten bivirkninger.	A	1a
Hos pasienter med meget høy samlet kardiovaskulær risiko inkludert diabetikere, kan lavere behandlingsmål vurderes (**).	B	1b
Hos pasienter med bivirkning av statinbehandling kan doseringen reduseres til den dosen som tolereres for å unngå seponering.	B	2b
Pasienter som behandles med et statin når de får hjerneinfarkt eller TIA, bør kontinuere behandlingen gjennom hele akuttfasen (ev. via nasogastrisk sonde).	A	1b

* se også kapittel 4.8 Levevaner

** se kapittel 4.6 Behandling ved diabetes mellitus

Figure B.4: AGR module without Bootstrap styling

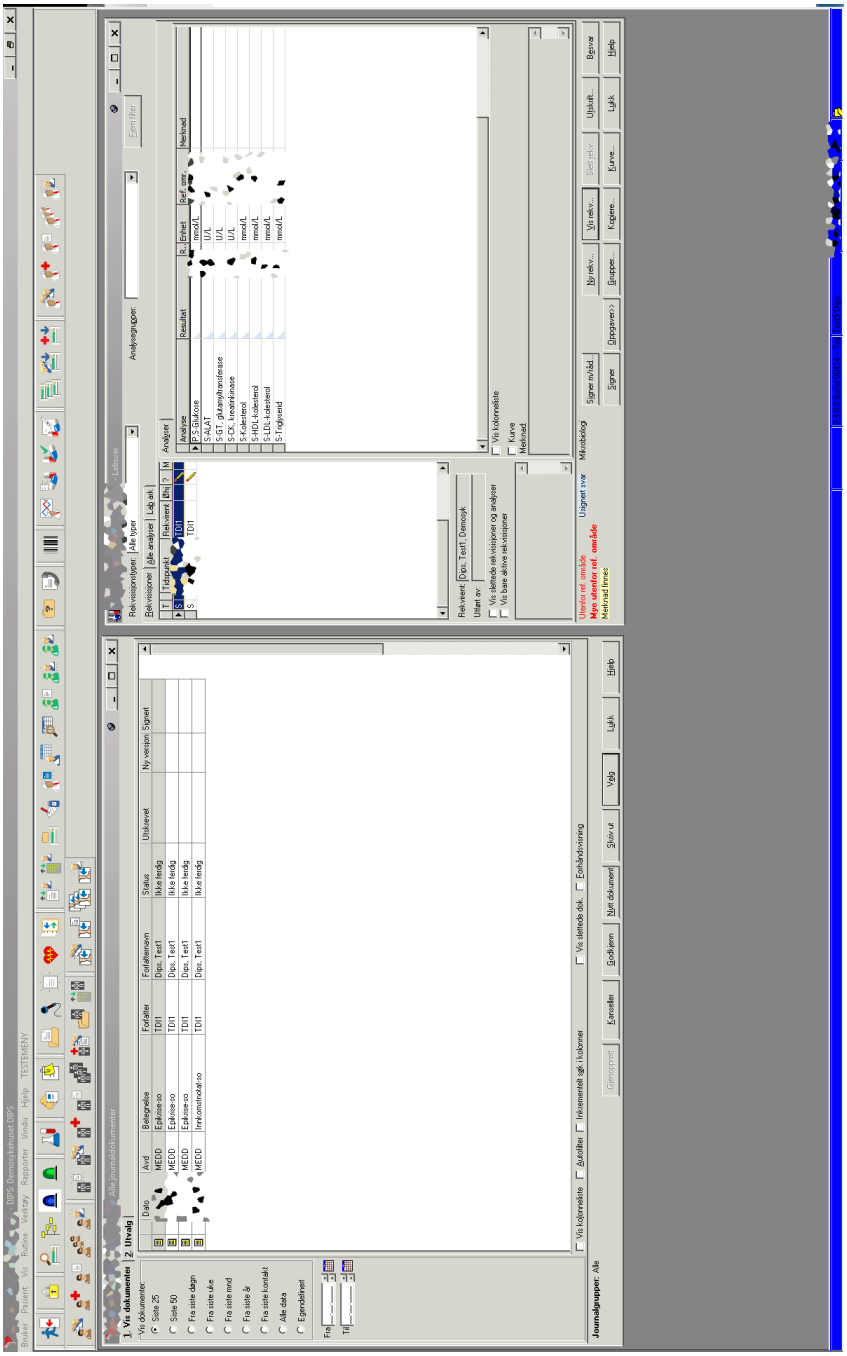


Figure B.5: Screenshot of DIPS

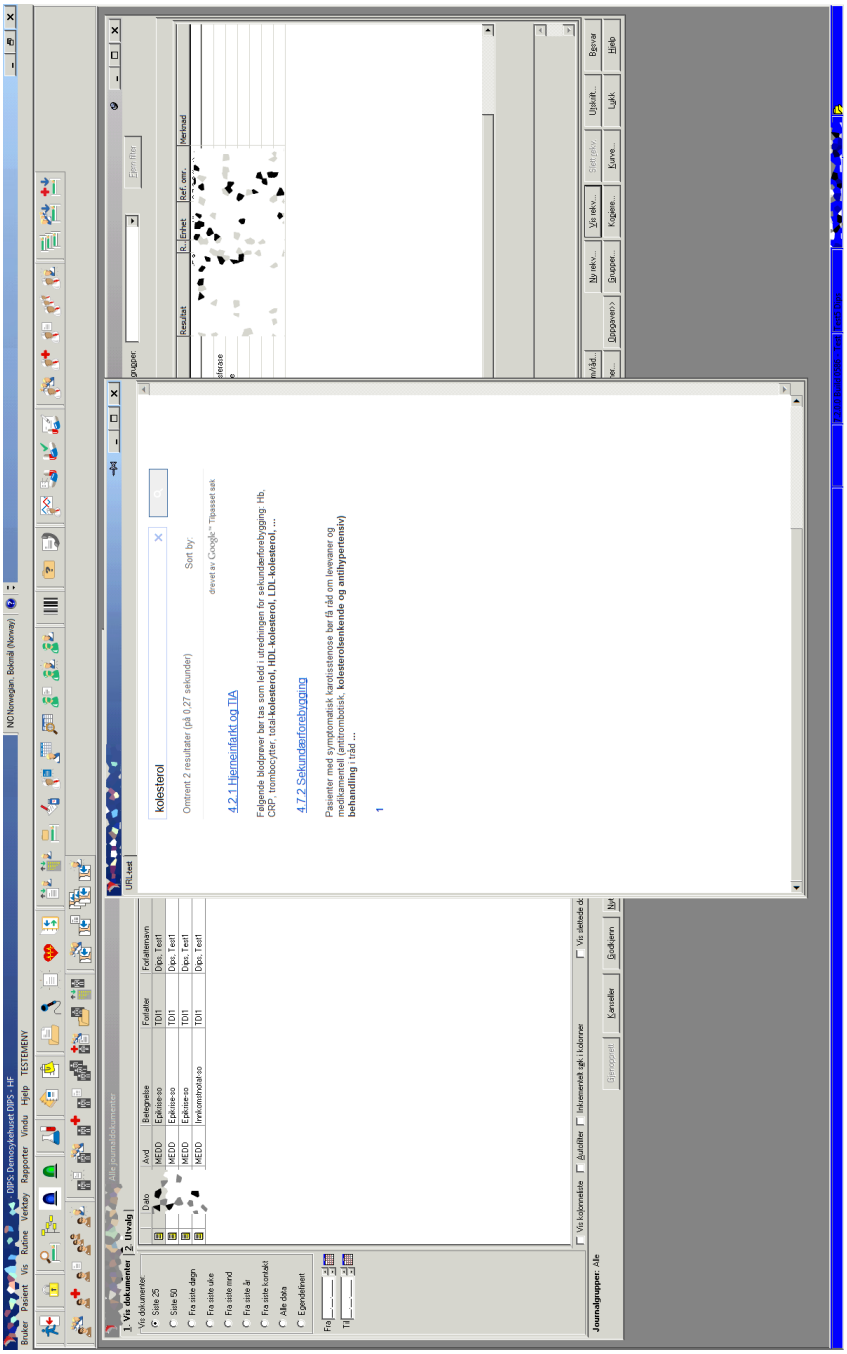


Figure B.6: Screenshot of DIPS with GAS

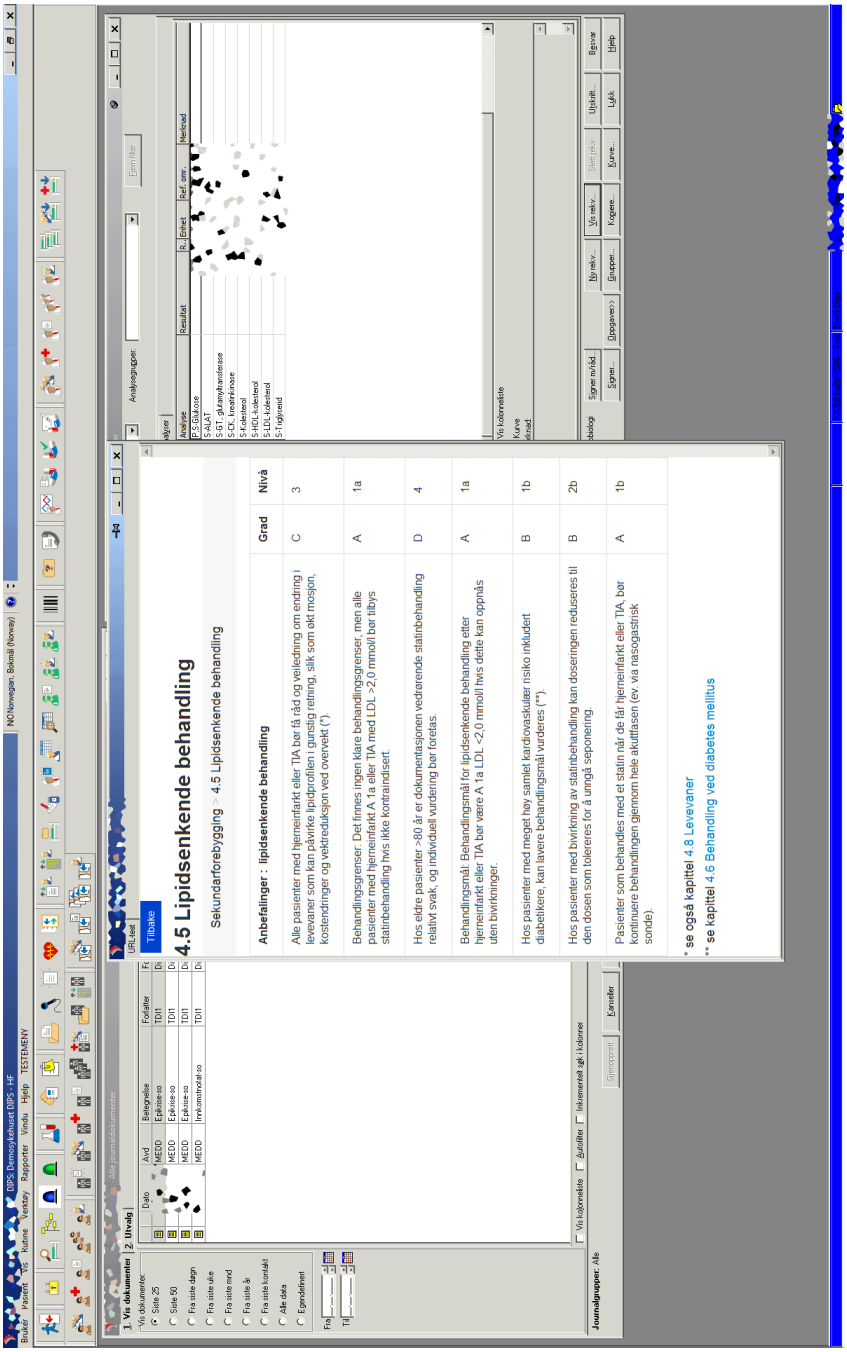


Figure B.7: Guideline for cholesterol lowering treatment in GAS

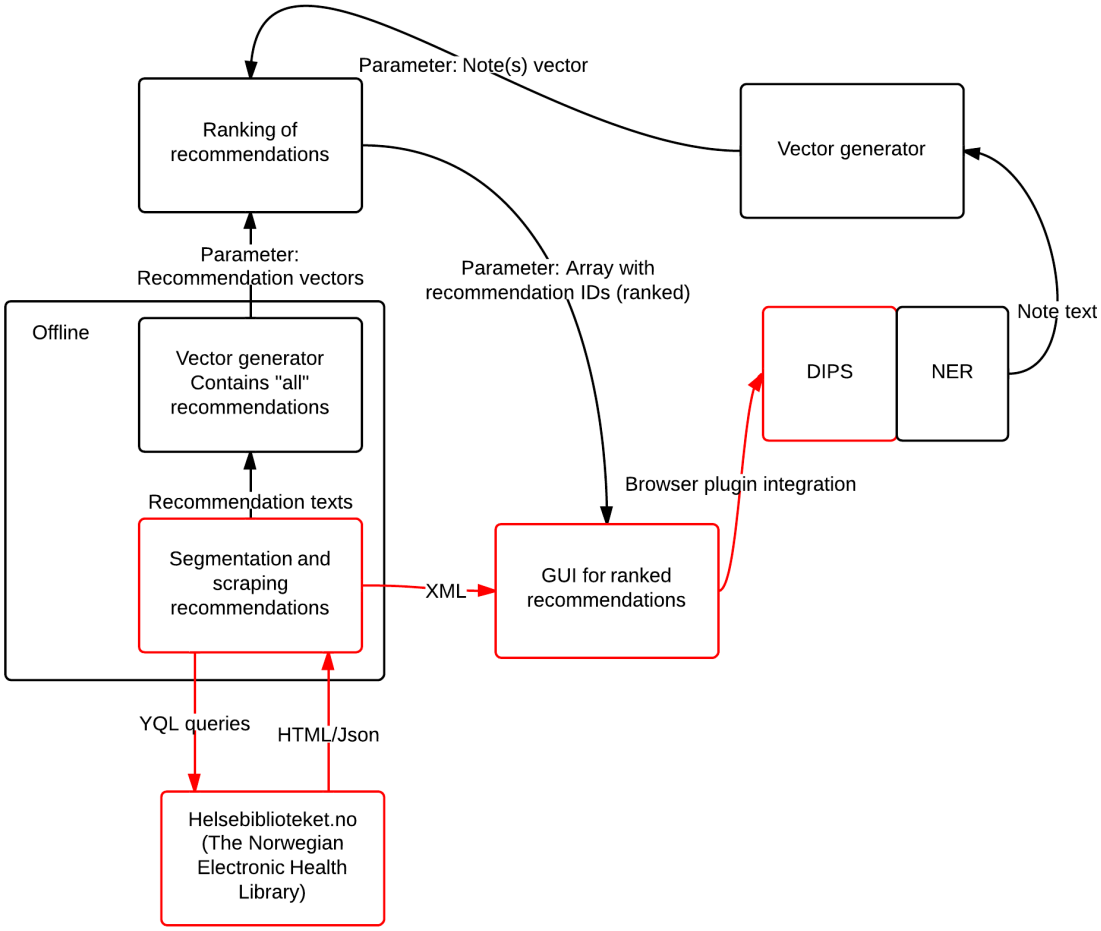


Figure B.8: Overview over this and adjacent projects

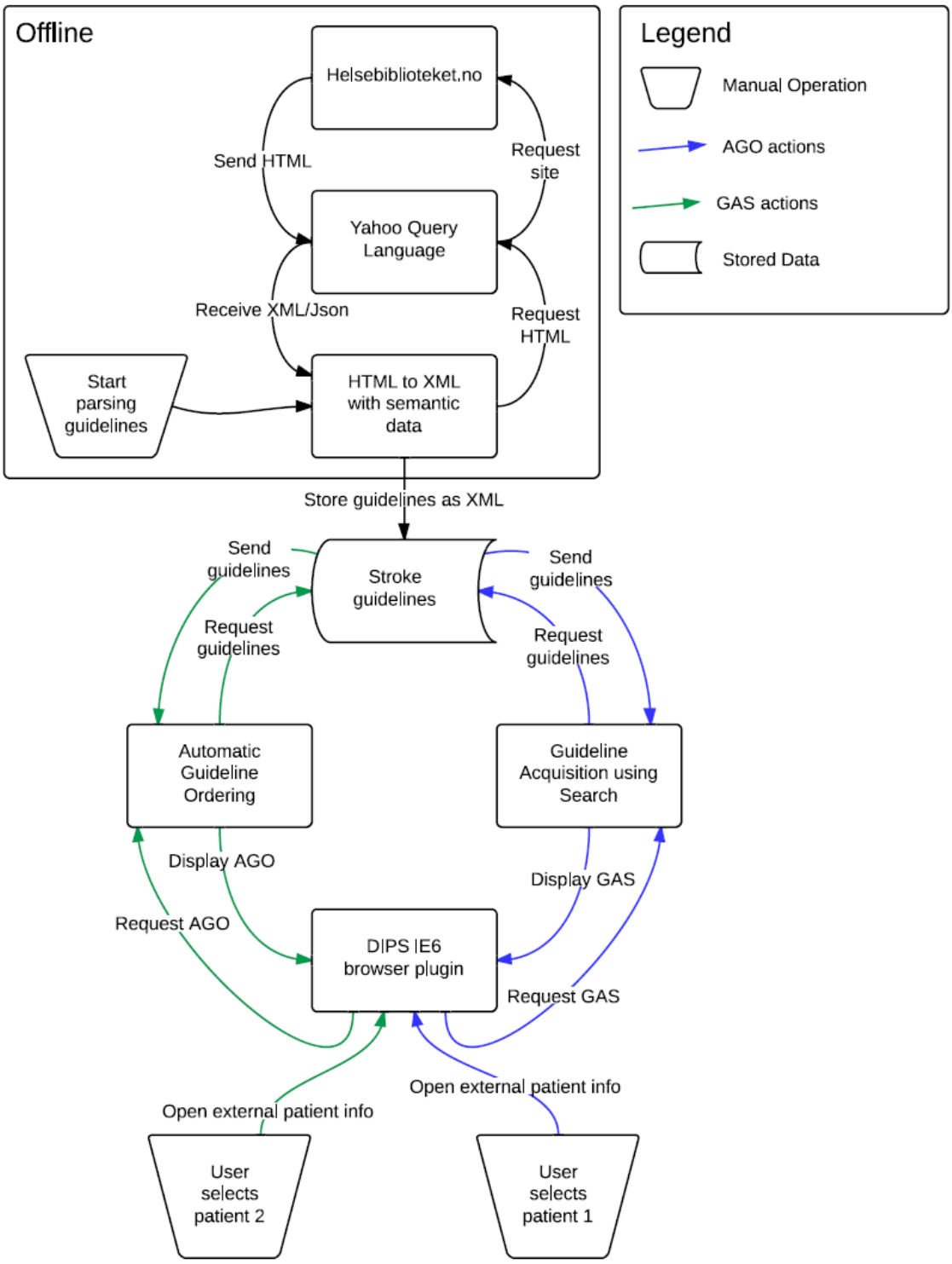


Figure B.9: Overview over GAS, AGR and DIPS