



CHALMERS



GÖTEBORGS UNIVERSITET



# Utveckling av ett AI-drivet stödsystem för navigering av administrativa system

Examensarbete inom högskoleprogrammet Datateknik

JENNY MÅNSSON  
CORNELIA SWARTLING

**INSTITUTIONEN FÖR DATA- OCH INFORMATIONSTEKNIK**

CHALMERS TEKNISKA HÖGSKOLA  
Göteborg 2024  
[www.chalmers.se](http://www.chalmers.se)



EXAMENSARBETE 2024

Utveckling av ett AI-drivet stödsystem för  
navigering av administrativa system

JENNY MÅNSSON  
CORNELIA SWARTLING



GÖTEBORGS  
UNIVERSITET

---



**CHALMERS**

Institutionen för data- och informationsteknik  
CHALMERS TEKNISKA HÖGSKOLA  
GÖTEBORGS UNIVERSITET  
Göteborg 2024

Utveckling av ett AI-drivet stödsystem för navigering av administrativa system  
JENNY MÅNSSON  
CORNELIA SWARTLING

© JENNY MÅNSSON, CORNELIA SWARTLING, 2024.

Handledare: Sakib Sistik, Institutionen för data- och informationsteknik  
Examinator: Jonas Almström Duregård, Institutionen för data- och informations-  
teknik

Examensarbete 2024  
Institutionen för data- och informationsteknik  
Chalmers Tekniska Högskola  
SE-412 96 Göteborg  
Telefon +46 31 772 1000

Omslagsbild: AI-genererad bild av en dator. Källa: Canva AI [1].

Skriven i L<sup>A</sup>T<sub>E</sub>X  
Göteborg 2024

Utveckling av ett AI-drivet stödsystem för navigering av administrativa system  
JENNY MÅNSSON & CORNELIA SWARTLING  
Institutionen för data- och informationsteknik  
Chalmers Tekniska Högskola  
Göteborgs Universitet

## Sammanfattning

Detta examensarbete fokuserar på att utveckla ett AI-drivet stödsystem för att förbättra navigeringen av Chalmers tekniska högskolas intranät. Målet är att underlätta för användare att snabbt hitta relevant information, såsom manualer, länkar till externa webbsidor och kontaktpersoner.

Metoden innefattade utveckling och implementering av olika AI-modeller för att skapa ett system kapabelt att hantera både svenska och engelska. Flera modeller testades och jämfördes för att identifiera den mest optimala för uppgiften. Applikationen utvärderades med hjälp av verklig data från Chalmers ekonomiavdelning för att mäta dess prestanda och noggrannhet.

Resultatet visar att KBLab/sentence-bert-swedish-cased var den mest effektiva modellen, vilket möjliggjorde en korrekt och precis sökning. Stödsystemet visade sig vara kapabelt att effektivt hitta relevant information från dess databas baserat på användarens söktermer och därigenom förbättra användarupplevelsen och effektivisera arbetsflödet.

Slutsatsen är att det utvecklade stödsystemet avsevärt kan förbättra informations-sökningen på Chalmers intranät. Framtida versioner av systemet kan utökas med förbättrade AI-modeller och träningsalgoritmer, samt förbättrade säkerhetsåtgärder och ett mer intuitivt användargränssnitt. Examensarbetet lägger därmed en stabil grund för fortsatt utveckling och implementering av avancerade stödsystem inom liknande organisationer.

Nyckelord: artificiell intelligens, AI, maskininlärning, ML, webbspindel, sentence transformer, stödsystem, sökmotor, semantisk sökning, cosine similarity.



Utveckling av ett AI-drivet stödsystem för navigering av administrativa system  
JENNY MÅNSSON & CORNELIA SWARTLING  
Institutionen för data- och informationsteknik  
Chalmers Tekniska Högskola  
Göteborgs Universitet

## Abstract

This thesis focuses on developing an AI-driven support system to improve the navigation of the intranet at Chalmers University of Technology. The aim is to facilitate users' ability to quickly find relevant information, such as manuals, links to external websites, and contact information.

The method included the development and implementation of different AI models to create a system capable of handling both Swedish and English. Several models were tested and compared to identify the most optimal one for the task. The application was evaluated with real data from the economy department of Chalmers, to measure its performance and accuracy.

The results exhibit that KBLab/sentence-bert-swedish-cased was the most efficient model, making a correct and precise search possible. The support system displayed its capability to effectively find relevant information from its database based on users' search terms, thereby improving the user experience and streamlining the workflow.

The conclusion is that the developed support system could improve searching for information on Chalmers' intranet significantly. Future versions of the system could be further developed with improved AI models and training algorithms, as well as with improved security measures and a more intuitive user interface. Thereby, this thesis lays a solid foundation for further development and implementation of advanced support systems within similar organizations.

This report is written in Swedish.

Keywords: artificial intelligence, AI, machine learning, ML, web spider, sentence transformer, support system, search engine, semantic search, cosine similarity.



## Förord

Denna rapport ger läsaren en inblick i hur ett stödsystem kan implementeras och hur det kan användas för att effektivisera arbetsflöden. Dessutom diskuteras värdefulla etiska aspekter som är viktiga att beakta vid implementering och användning av ett sådant system.

Tack till Sanna Erling-Almér och Chalmers ekonomiavdelning som har gjort detta examensarbete möjligt och sett till att vi fått det material vi behöver, samt gett oss värdefulla insikter i hur ett system hade kunnat effektivisera arbetsflödet.

Även ett stort tack till vår handledare Sakib Sisteck som har väglett oss genom detta examensarbete med sitt stöd och sin expertis.

Jenny Månsson och Cornelia Swartling, Göteborg, maj 2024



# Terminologi

all-MiniLM-L6-v2	Sentence transformer, betecknad ST1.
all-mpnet-base-v2	Sentence transformer med större vektorrum, betecknad ST2.
Artificiell Intelligens	Teknik ämnad för att efterlikna mänsklig intelligens.
Cosine similarity	Mätvärdet för semantisk likhet.
Django	Python-webbframverk för snabb utveckling.
Feature extraction	Identifiering och extraktion av relevanta egenskaper från data.
GPT-SW3	Generativ språkmodell för svenska från AI-Sweden.
KBLab/sentence- bert-swedish-cased	Sentence transformer tränad på engelska och svenska, betecknad ST3.
Maskininlärning	Underkategori av AI där algoritmen lär sig från data.
Python	Programmeringsspråk som projektet är utvecklat i.
Sentence transformer	Modell som transformerar meningar till vektorrepresentationer.
Tröskel	Ett cosine similarity-värde som avgör gränsen för dokumentrelevans.
Vektorrum	Matematiskt rum där data representeras som vektorer.
Webbspindel	Programvara som automatiskt samlar in data från webbsidor.



# Innehåll

<b>Akronymer</b>	<b>xi</b>
<b>Figurer</b>	<b>xv</b>
<b>Tabeller</b>	<b>xvii</b>
<b>1 Inledning</b>	<b>1</b>
1.1 Bakgrund . . . . .	1
1.2 Syfte . . . . .	1
1.3 Mål . . . . .	1
1.4 Avgränsningar . . . . .	2
<b>2 Teknisk Bakgrund</b>	<b>3</b>
2.1 Python . . . . .	3
2.2 Artificiell Intelligens . . . . .	3
2.2.1 Feature extraction . . . . .	4
2.2.2 GPT-SW3 . . . . .	4
2.2.3 Sentence Transformers . . . . .	4
2.2.3.1 Cosine similarity . . . . .	5
2.3 Webbspindel och Scrapy . . . . .	5
<b>3 Metod</b>	<b>7</b>
3.1 Utveckling av sökfunktion . . . . .	7
3.2 Utveckling av webbspindel . . . . .	8
3.3 Utveckling av webbsidan . . . . .	8
3.4 Utveckling av databas . . . . .	9
3.5 PDF-hantering . . . . .	9
3.6 Evaluering av applikation . . . . .	10
<b>4 Resultat</b>	<b>11</b>
4.1 Applikation . . . . .	11
4.2 Tester av applikation . . . . .	12
<b>5 Diskussion</b>	<b>17</b>
5.1 Metoddiskussion . . . . .	17
5.2 Val av Sentence Transformer . . . . .	18
5.3 Jämförelse med andra sökmotorer . . . . .	19

## Innehåll

---

5.4	Förbättringsområden och vidareutveckling . . . . .	20
5.5	Etik och miljö . . . . .	20
5.5.1	Etik . . . . .	20
5.5.2	Miljö . . . . .	21
<b>6</b>	<b>Slutsats</b>	<b>23</b>
	<b>Bibliography</b>	<b>25</b>
<b>A</b>	<b>Appendix 1</b>	<b>I</b>

# Figurer

2.1	Djangos uppbyggnad . . . . .	3
2.2	Representation av cosine similarity och vektorers relation till dess värde. . . . .	5
3.1	En skärmbild av webbsidans första prototyp, före och efter sökning “vilka språk läser man i skolan?”. . . . .	9
4.1	En skärmbild av den färdiga webbsidan, före sökning . . . . .	11
4.2	En skärmbild av den färdiga webbsidan, före sökning . . . . .	12



# Tabeller

4.1	Filstorlek och tid för uppladdning av PDF-filer och träning för de olika sentence transformers (ST) . . . . .	13
4.2	Storlek och tid för uppladdning av webbsidor och träning för de olika sentence transformers (ST) . . . . .	13
4.3	Cosine similarity-värden och uppmätt tid från sökningen “redovisning” för de olika sentence transformers . . . . .	14
4.4	Cosine similarity-värden och uppmätt tid från sökningen “Hur gör man ett bokslut?” för ST3 . . . . .	15
4.5	Cosine similarity-värden och uppmätt tid från sökningen “fotboll” för ST3. . . . .	16
A.1	Filstorlek och tid för uppladdning av PDF-filer och träning för de olika sentence transformers (ST) . . . . .	I
A.2	Storlek och tid för uppladdning av webbsidor och träning för de olika sentence transformers (ST) . . . . .	I
A.3	Cosine similarity-värden och uppmätt tid från sökningen “redovisning” för de olika sentence transformers . . . . .	II
A.4	Cosine similarity-värden och uppmätt tid från sökningen “Hur gör man ett bokslut?” för ST3 . . . . .	III
A.5	Cosine similarity-värden och uppmätt tid från sökningen “fotboll” för ST3. . . . .	IV



# 1

## Inledning

I dagens digitaliserade värld är effektiv informationssökning avgörande för att underlätta arbetsflöden och förbättra produktivitet. Med den ständigt växande digitala transformationen finns det fler möjligheter än någonsin att göra sökandet efter information och data mer och effektivt och lättillgängligt. Administrativa system kan vara komplexa och svåra att navigera i, och ibland leda en runt i cirklar. Genom att utveckla en AI-driven sökmotor som kan ta fram relevanta dokument, lathundar eller länkar till hemsidor kan ett smidigare system uppnås, för att underlätta för användare.

### 1.1 Bakgrund

Chalmers tekniska högskola har behovet att uppdatera hur deras intranät är uppbyggt. En anställd på Chalmers spenderar tid av sin arbetsdag bara på att försöka hitta rätt dokument eller e-tjänst i en djungel av länkar till olika sidor. Det befintliga söksystemet gynnar endast de som är välbekanta med det, och de länkar som finns nedsparade lokalt riskerar att byta plats i systemet och sluta fungera. Därmed hamnar den anställde tillbaka i sökmotorn där än mer tid förloras, vilket bygger upp ett behov av ett bättre, optimerat system.

### 1.2 Syfte

Syftet med detta examensarbete är att utveckla ett AI-drivet stödsystem som kan underlätta för Chalmersanställda. Det ska guida användarna genom delar av Chalmers intranät, specifikt ekonomi- och administrationsdelen. Systemet ska kunna tolka sökfraser och frågor och därefter presentera relevant information. Detta ska möjliggöra för en effektiv och mer precis navigering i organisationens databas.

### 1.3 Mål

- Utforska och definiera användarbehov angående system för att söka i Chalmers intranät
- Utveckla en prototyp av ett system som använder sig av AI för att ta fram relevanta sökresultat
- Testa och utvärdera prototypen på Chalmers data

- I mån av tid implementera fler möjligheter för att kunna söka utanför Chalmers databas, såsom Skatteverket och Bolagsverket

### 1.4 Avgränsningar

- Initialt kommer systemet att utvecklas för att hantera frågor och sökningar inom en lokal databas. Detta kan senare komma att utvecklas till att innefatta organisationens data och system. Projektet kommer då att fokusera på att förbättra tillgången till information inom den specifika organisationen och dess interna system. Det kommer inte att omfatta externa informationskällor som inte är relevanta för organisationens verksamhet.
- Integrering med externa system kommer att betraktas som en möjlig framtida potential och inte som en omedelbar prioritet.
- Utvecklingen kommer att fokusera på att bygga en AI-algoritm för att effektivt analysera och förstå användarfrågor samt att bygga en användarvänlig applikation för att tillhandahålla vägledning. Djupgående tekniska lösningar som kräver betydande resurser eller infrastruktur-förändringar kommer inte att utforskas i denna fas.

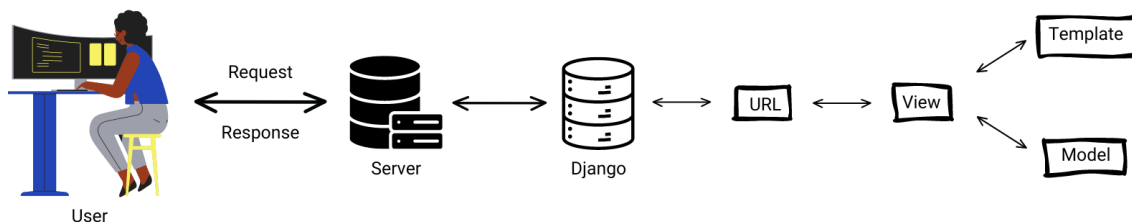
# 2

## Teknisk Bakgrund

I följande avsnitt beskrivs de tekniska koncept och verktyg som används i projektet. Fokus ligger på sentence transformers och webbspindlar, deras funktion och hur de tillämpas i systemet. Dessa två komponenter är essentiella för att möjliggöra korrekt och effektiv informationssökning. Vidare presenteras hur dessa teknologier har implementerats för att uppnå projektets mål.

### 2.1 Python

Stödsystemet är främst utvecklat i Python, ett programmeringsspråk som är välkänt för sitt stöd för maskininlärning och dataanalys. Python erbjuder ett omfattande ekosystem av bibliotek och ramverk, såsom PyTorch och Scikit-learn, vilket gör språket till ett passande val för detta projekt [2, 3]. Ett specifikt ramverk som används i detta projekt är Django. Django är ett ramverk för att bygga webbapplikationer med Python, vilket ger möjlighet till smidig integration med övriga delar av projektet som också är skrivna i Python [4]. Detta ramverk innehåller allt som en utvecklare kan behöva vid ett projekt, vilket innebär att både databasen, back-end och webbapplikationen är implementerade i Python.



Figur 2.1: Djangos uppbyggnad

### 2.2 Artificiell Intelligens

Artificiell Intelligens (AI) har förmågan att simulera mänsklig intelligens och problemlösningsförmåga. Genom att analysera stora mängder data och använda träningsalgoritmer kan AI lära sig och göra mer exakta förutsägelser baserade på ny data. Förmågan att lära sig från data är känd som maskininlärning, en central komponent inom AI [5]. I detta stödsystem används AI för att tolka användarens

söktermer och extrahera relevant data från databasen för att hjälpa användaren att navigera. Dessa verktyg beskrivs mer i detalj i kapitlen nedan.

### 2.2.1 Feature extraction

Feature extraction (sv. funktionsutvinning) är processen att identifiera och extrahera relevanta egenskaper eller attribut från data, som sedan används som indata för maskininlärningsalgoritmer. Denna process är kritisk eftersom den påverkar AI-modellens förmåga att förstå och lära sig från data [6]. Då det svenska namnet inte är välkänt, används det engelska ordet i rapporten.

### 2.2.2 GPT-SW3

AI Swedens modell *GPT-SW3* är en anpassad version av Generative Pre-trained Transformer (GPT). Denna modell är optimerad för att hantera och generera svar baserat på instruktioner genom att använda Transformer-arkitekturen. Modellen är tränad på ett dataset innehållande tokens på svenska, norska, danska, isländska, engelska samt programmeringskod, och har förtränts med hjälp av Causal Language Modeling (CLM). CLM innebär att modellen tränas för att förutsäga nästa ord i en sekvens, vilket hjälper till att utveckla en djup förståelse för språkstruktur och semantiska sammanhang [7, 8]. Genom denna modell kan applikationen tolka användarens söktermer och generera relevanta svar, vilket gör den särskilt lämplig för ett stödsystem som detta.

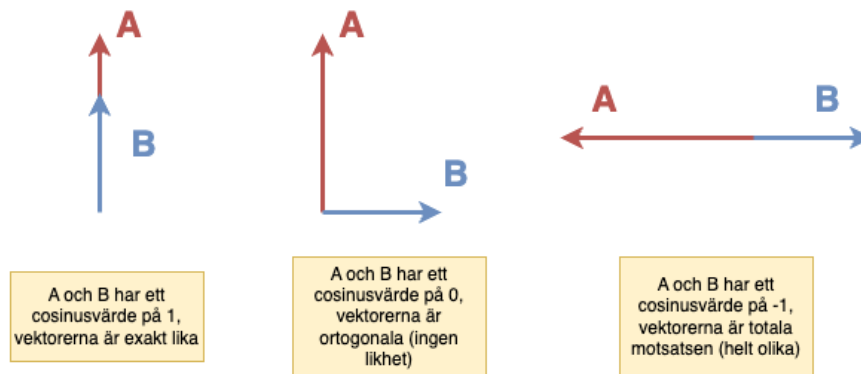
### 2.2.3 Sentence Transformers

Sentence transformers (sv. meningstransformationer) är ett bibliotek som erbjuder tusentals förtränsade modeller för att utföra uppgifter på olika typer av data, inklusive text, bild, och ljud. Inom textbehandling kan modellerna användas för uppgifter som textklassificering, informationsextraktion samt frågor och svar [9]. I detta projekt användes modellerna *all-MiniLM-L6-v2* [10], *all-mpnet-base-v2* [11] och *KBLab/sentence-bert-swedish-cased* [12, 13]. Dessa modeller betecknas även som ST1, ST2 respektive ST3 i denna rapport. Samtliga modeller kan användas för semantisk sökning, vilket betyder att de kan förstå sambanden mellan ord och därmed extrahera resultat från databasen som är relevanta till sökningen och inte bara matcha orden i söktermen [14].

Sentence transformer-modellerna kartlägger texter till ett vektorrum med ett specifikt antal dimensioner. Antalet dimensioner påverkar modellens kapacitet att representera semantiska likheter och skillnader mellan texter, vilket är en viktig egenskap i detta projekt [15]. I denna rapport kommer de att benämnas som sentence transformers eftersom det svenska ordet inte är lika välkänt. Även benämningen likhetsvärde kan förekomma.

### 2.2.3.1 Cosine similarity

Cosine similarity (sv. cosinuslikhet) är ett matematiskt mått på likhet mellan två vektorer. Det beräknar cosinusvärdet av vinkeln mellan två vektorer i ett vektorrum. Värdet sträcker sig från -1 till 1, där ett värde närmare 1 indikerar större likhet mellan vektorerna, som illustreras i figur 2.2 [16, 17].



**Figur 2.2:** Representation av cosine similarity och vektorers relation till dess värde.

Efter att sentence transformer-modellerna kartlagt texterna till ett vektorrum, kan den semantiska likheten kalkyleras mellan dessa vektorer med hjälp av cosine similarity. Genom att jämföra vektorn för en sökfras och vektorerna för dokumenten i databasen, kan systemet rangordna dokumenten baserat på deras semantiska likhet med sökfrasen.

I detta projekt används cosine similarity för att hitta semantiska likheter mellan texter eftersom det effektivt mäter likheten oberoende av textens längd. Detta beror på att måttet är baserat på vinkeln mellan vektorerna, inte deras storlek. Detta är särskilt viktigt då sökfrasen och dokumenten varierar i storlek. Den engelska termen, cosine similarity, kommer att användas i rapporten eftersom denna är mer välkänd än den svenska benämningen cosinuslikhet.

## 2.3 Webbspindel och Scrapy

För att genomföra en sökning på webbsidor och dess relaterade länkar krävs en spindel (även kallad web crawler). I denna rapport benämns den som en webbspindel. Denna webbspindel är utformad för att hämta webbsidor och identifiera nya sidor genom att systematiskt följa hyperlänkar som finns på sidorna. Med hjälp av webbspindeltekniken kan omfattande mängder med data relaterade till sökningen samlas in [18].

Scrapy är ett ramverk för webbskrapning som används för att utveckla webbspindlar, som navigerar och extraherar data från hemsidor. Ramverket är skrivet i Python och tillhandahåller funktioner för att hantera nedladdning av sidor, datahämtning, och lagring av skrapad information. [19]



# 3

## Metod

Projektet startade med ett möte med den externa handledaren Sanna Erling-Almér, administrativ chef på institutionen för Data- och informationsteknik. Detta möte resulterade i en bredare förståelse för båda parter av vad projektet förväntades resultera i. Därefter påbörjades projektets praktiska delar genom att en enklare skiss över modellens uppbyggnad skapades.

### 3.1 Utveckling av sökfunktion

Utvecklingen av uppladdnings- och sökfunktionen påbörjades först eftersom detta utgjorde en central del av projektet. Denna del förväntades ta längst tid och stort fokus låg på att få en uppfattning om tidsåtgången tidigt i processen genom kartläggning och planering. Samtidigt genomfördes det en teknisk fördjupning i flera områden, såsom webbspindlar och Django, eftersom kunskapen om dem var begränsad. En djupare förståelse för dessa områden skulle gynna projektets framgång, och prioriterades tidigt i processen.

Datan från Chalmers som skulle användas till att träna modellen på fick fördröjningar och det beslutades att implementera en generell applikation för att extrahera text från webbsidors HTML-filer och träna modellen på den datan istället. Under tiden undersöktes vilken förtränad modell som skulle vara optimal i applikationen. Genom att använda förtränade modeller kunde utvecklingsprocessen snabbas upp och sökresultatens kvalitet och korrekthet förbättras. Eftersom dessa modeller redan tränats på stora mängder data hade de en god grundförståelse för språkliga mönster och sammanhang. Den modell som först testades var AI Swedens “GPT-SW3”, en generativ språkmodell för de nordiska språken, främst svenska [7, 8]. Med generativ menas att modellen kan komplettera text, vilket kan vara en fördel vid sökning och informationsextrahering. Dessutom behövdes en kompletterande modell eftersom målet var att hitta relevanta dokument i databasen till användarens sökning. Därför beslutades det att använda GPT-SW3 för feature extraction och en annan modell för att hitta likheter.

För att hantera semantisk sökning valdes den förtränade modellen *all-MiniLM-L6-v2* [10]. Med denna modell kunde en sökmotor skapas som extraherade relevanta dokument från söktermer. Modellen behandlade dokumenten i databasen genom att kartlägga texterna till ett dimensionellt vektorrum. Nackdelen med denna modell var att den endast var tränad på engelsk data, vilket betyder att den inte var optimal

för en applikation som även ska behandla svenska. Den förtränade modellen kartlade även söktermerna till ett vektorrum för att hitta semantiska likheter mellan söktermerna och informationen i databasen. Genom att använda cosine similarity-värdet för respektive likheter kan dessa sökresultat rangordnas. En tröskel på 0.4 sattes för likhetsvärdet så att inte irrelevant data skulle klassas som ett relevant sökresultat. Detta innebar att endast dokument som hade ett likhetsvärde över 0.4 extraherades från databasen och visades som ett sökresultat. Den temporära databasen bestod av lokala HTML-filer från olika Wikipedia-sidor för att kunna testa programmet. När en enkel prototyp av applikationen var färdigställd, med den grundläggande sökfunktionen, var nästa steg att utveckla en webbspindel för att dynamiskt kunna hämta och extrahera data från internet.

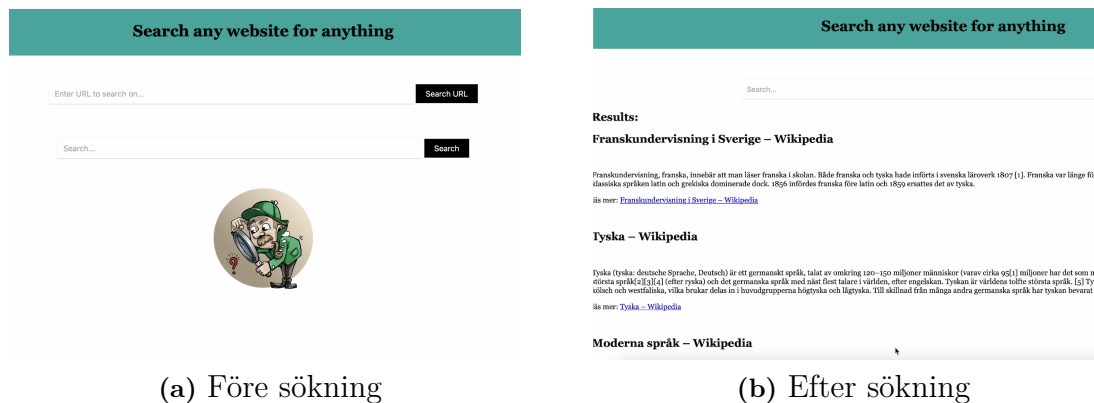
## 3.2 Utveckling av webbspindel

Webbspindel som användes i projektet var uppbyggd av Scrapy. Genom att ärva från Scrapy's webbspindel-klass definierades en egen klass specifikt för detta projekt. Denna klass fungerade genom att användaren angav en länk som utgjorde grunden för sökningen. Därefter kröp webbspindeln genom hyperlänkarna på den angivna sidan och samlade in relaterad data. Denna data användes sedan för att välja ut de mest relevanta sidorna från sökningen [18]. För denna prototyp användes ett djup på 1, vilket innebar att endast hyperlänkarna på den angivna sidan behandlades. Detta val gjordes baserat på de begränsningar i tid och datorkraft som fanns tillgängliga. Applikationen behandlade nu den angivna webbsidan och dess närliggande sidor för att kunna extrahera relevanta sidor för söktermerna som användaren angav. Eftersom det inte fanns ett användargränssnitt gjordes testerna i terminalen. Detta innebar att nästa steg var implementationen av en webbsida, då det fanns en prototyp som fungerade för alla typer av webbsidor.

## 3.3 Utveckling av webbsidan

I användargränssnittet till applikationen integrerades webbspindeln och sökfunktionen som utvecklades i föregående faser i projektet. Målet var att skapa en lättanvänd plattform för att underlätta för användaren. Eftersom de tidigare utvecklade funktionerna var implementerade i Python, användes ramverket Django. Detta underlättade utvecklandet av webbsidan då Django innehåller verktyg och bibliotek för att bygga webbapplikationer i Python [4]. Till en början implementerades endast en enkel webbsida, där användaren kunde ange en URL till en webbsida som grund för sökningen. Efter att AI-modellen hade processat all data från webbspindeln kunde användaren söka på datan genom att ange en sökfras. Som figur 3.1 visar fungerade webbsidan som tänkt. Här användes Wikipedia-sidan "Skolämnena i Sverige"[20] där sökfrasen var "vilka språk läser man i skolan?". Figur 3.1b) visar sökresultatet, vilket påvisar att AI-modellen har lyckats extrahera relevanta webbsidor för sökningen.

Parallellt med implementationen av webbsidan pågick också forskning efter en optimal sentence transformer som kunde hantera både svenska och engelska, för att



(a) Före sökning

(b) Efter sökning

**Figur 3.1:** En skärmbild av webbsidans första prototyp, före och efter sökning “vilka språk läser man i skolan?”.

säkerställa ett så effektivt och användbart stödsystemet som möjligt. Att effektivisera AI-modellen var även något som pågick under hela projektets gång.

### 3.4 Utveckling av databas

När webbsidan var implementerad var nästa steg att skapa en databas som kunde lagra alla webbsidor och dess data som webbspindeln extraherade. Detta för att undvika att behöva hämta all data på nytt varje gång användaren ville göra en ny sökning. Det skulle även vara ineffektivt och tidskrävande att behöva träna om AI-modellen på samma data vid varje ny sökning. Målet var också att ge användaren möjlighet att anpassa sin databas efter egna behov och kunna hämta information från olika hemsidor, oavsett om de är relaterade genom hyperlänkar eller inte. Ett exempel på detta behov kom från en insikt som handledaren Sanna gav: Chalmers anställda på ekonomiavdelningen kan behöva använda sig av både Skatteverket [21] och Bolagsverket [22] i sina arbetsuppgifter. Detta gjorde det optimalt för användarvänligheten att implementera en databas. Eftersom Django-ramverket även har inbyggt stöd för dataintegration, användes det för att enkelt konfigurera databasen, vilket möjliggjorde effektiv lagring och hämtning av data.

I denna fas av projektet kunde användaren lägga till webbsidor till databasen, där de lagrades tillsammans med sin URL, vilket gjorde dem unika i databasen. Även den tränade AI-modellen lagrades för att spara tid vid framtida sökningar. Med hjälp av databasen gjordes det möjligt för användaren att även lägga in flera olika hemsidor och söka över ett större utbud av information och därmed få ett större utbud av relevant data för sökningen.

### 3.5 PDF-hantering

När databasen var på plats och Chalmers data blev tillgänglig, övergick projektet till den femte fasen, implementering av PDF-hantering. Detta möjliggjorde för Chal-

mers anställda att ladda upp redan existerande lathundar och annan dokumentation för att behålla samma sökbank som tidigare, samtidigt som den kunde utökas med extern information från exempelvis Skatteverket, enligt handledaren Sannas önskemål. Innehållet i PDF-filerna extraherades och lagrades tillsammans med relevant information i databasen. Vid sökningar hämtades både relevanta hemsidor och PDF-filer, vilket bidrog till en mer optimal sökupplevelse.

## 3.6 Evaluering av applikation

I denna slutliga fas testades två andra sentence transformers: *all-mpnet-base-v2* [11] (ST2) och *KBLab/sentence-bert-swedish-cased* (ST3) [12]. ST2 var tränad på exakt samma data som den redan använda transformern *all-MiniLM-L6-v2* (ST1) [10], men kartlade texter till ett 768-dimensionellt vektorrum istället för ett 384-dimensionellt vektorrum. Eftersom antalet dimensioner påverkar modellens kapacitet att representera semantiska likheter, ansågs detta vara ett värdefullt test. ST3 hade ST2 som lärarmodell men var dessutom tränad på svenska, vilket teoretiskt gjorde ST3 till en optimal modell.

Testerna som utfördes inkluderade mätning av uppladdningstider för PDF-filer och webbsidor, samt söktester för att fastställa hur väl stödsystemet kan tolka användarens söktermer och extrahera relevanta dokument från databasen. Eftersom de tre olika modellerna hade olika förutsättningar, såsom antal dimensioner och variation i språk, gav de olika cosine similarity-värden. Högre värden indikerar större semantisk likhet, vilket gör att de extraherade dokumenten kan rangordnas efter relevans i sökresultatet. För att fastställa relevans användes en tröskel för cosine similarity-värdet. Dokumenten som översteg denna tröskel ansågs relevanta och inkluderades i resultaten. Tröskeln justerades baserat på intuitivt tänkande och cosine similarity-värdet för att hitta en rimlig balans mellan relevanta och irrelevanta resultat.

Testresultaten gav insikt i hur stödsystemet presterade med de olika sentence transformermodellerna och med detta resultat kan slutsatser dras om vilken av modellerna som var mest optimal för systemet. Den modell som genererade de högsta cosine similarity-värdena för relevanta dokument och samtidigt upprätthöll rimliga svarstider, användes vidare i projektet.

Under denna fas gjordes också försök att optimera träningen av AI-modellen och att hantera data på ett effektivt sätt. En del buggar och små justeringar åtgärdades, och dessa kommer att diskuteras mer ingående i ett senare avsnitt av rapporten. Jämförelser med andra sökmotorer och liknande program påbörjades också för att utvärdera och verifiera applikationens prestanda.

# 4

## Resultat

I detta kapitel presenteras de resultat som projektet kom fram till.

### 4.1 Applikation

Projektet resulterade i en webbsida, som visas i figur 4.1 och 4.2, där användare kan lägga in webbsidor eller PDF-filer för att anpassa sin sökbank till sina behov. Filstorleken per uppladdning kan nå upp till maximalt 10 MB, medan webbsidorna bearbetas till ett djup på en nivå, vilket innebär att endast de webbsidor som är direkt länkade från den ursprungliga webbsidan kommer att behandlas. Datan från filerna eller webbsidorna som laddats upp behandlas av en AI-modell. Därefter kan användaren söka på termer eller frågor och få upp relevanta sökresultat baserat på dessa termer. Från dessa sökresultat kan användaren navigera vidare genom länkar till den efterfrågade informationen.

The screenshot displays a dark-themed web application interface. At the top, a dark blue header contains a small circular logo on the left and the text "Search any website or pdf for anything" in white. Below the header, the main content area is divided into two sections. The first section is titled "Input website address or select PDF file" and contains two input options: a text input field with the placeholder "Enter URL to search on..." and a "Search URL" button, and a file selection area with a "Choose Files" button, the text "no files selected", a "Max 10 MB" limit, and an "Upload PDF" button. The second section is titled "Input question or search phrases" and features a single text input field with the placeholder "Search..." and a "Search" button.

**Figur 4.1:** En skärmbild av den färdiga webbsidan, före sökning

Figur 4.1 visar slutprodukten av detta projekt före sökning. Här kan man se att samtliga funktioner har lagts till sedan första prototypen av webbsidan, exempelvis uppladdning av PDF-filer, samt ett mer användarvänligt utseende. Figur 4.2 visar sökresultaten för "Hur gör man ett bokslut?" för att illustrera att applikationen även kan ta emot frågor och inte bara enskilda termer. De två första sökresultaten är från Skatteverkets hemsida och en PDF-fil från Chalmers data. Vidare beskrivs de tester som genomfördes för att avgöra vilken av de testade sentence transformermodellerna som var mest optimal för applikationen.



Figur 4.2: En skärmbild av den färdiga webbsidan, före sökning

## 4.2 Tester av applikation

Dessa tester omfattade jämförelser av semantisk likhet och prestanda mellan de olika modellerna. För att bedöma prestandan av stödsystemet bedömdes testerna utifrån dessa kriterier:

- Uppladdning av PDF (tid).
- Uppladdning av webbsida (tid).
- Sökning (cosine similarity och söktid).

I det första testet utvärderades uppladdningstiden för PDF-filer. PDF-filerna från Chalmers användes för att göra testerna så relevanta som möjligt. För att hålla testerna koncisa, valdes cirka en tredjedel av filerna med varierande storlek. Även de tre olika sentence transformers testades för att få en bredare översikt och förståelse över hur de bearbetade datan. Resultaten av detta test visas i tabellen 4.1, där kolumnerna ST representerar de olika sentence transformers som testades: *all-MiniLM-L6-v2* (ST1) [10], *all-mpnet-base-v2* (ST2) [11] och *KBLab/sentence-bert-swedish-cased* (ST3) [12].

Som resultatet visar i tabellen 4.1 finns det inte en tydlig parallell mellan storleken på filen och den tidsåtgång som krävdes för att ladda upp den och träna datan med AI-modellen. Av den anledningen beslutades det att inte fortsätta med ytterligare tester av denna typ på PDF-filer. Dessutom observerades det att sentence transformer ST1 var den snabbaste när det kom till bearbetning av datan i filerna. Därefter utfördes ett nytt test. Liknande det första testet skulle uppladdningen av hemsidor utvärderas.

Filnamn	Filstorlek	ST1 (s)	ST2 (s)	ST3 (s)
Redovisningens syfte & principer	278 KB	29.51	40.95	42.24
Särskilda skatt- och momsregler	481 KB	31.86	45.63	47.69
Hantera fakturor – vanliga konto...	511 KB	28.31	47.26	51.82
Tidplan ekonomiarbete	844 KB	31.14	40.09	42.49
Kontoplan & objektsplan med info...	1,5 MB	32.84	51.35	51.36
Så här gör du inköp	2 MB	30.92	47.40	48.85
Kontaktpersoner ekonomistöd	5,3 MB	24.43	43.92	45.66
Projektkalkyl	5,8 MB	36.49	49.71	49.98
Genomsnittlig tid (s)		30.69	45.79	47.51

**Tabell 4.1:** Filstorlek och tid för uppladdning av PDF-filer och träning för de olika sentence transformers (ST)

Titel på hemsida	Storlek	ST1 (min)	ST2 (min)	ST3 (min)
Deklarera för föreningar	113 KB	3.68	3.69	9.18
E-tjänster och blanketter för företag	118 KB	3.91	3.99	10.03
Avdragsrätt, fakturering och redovisning	125 KB	5.55	5.41	13.42
Moms - Företag och organisationer	128 KB	6.23	4.33	10.73
Inkomst av näringsverksamhet	140 KB	6.98	6.95	17.31
Du är medborgare i Storbritannien med...	162 KB	5.83	5.90	14.90
Ränteutgifter	176 KB	0.14	6.12	16.36
Svar på vanliga frågor - deklaration...	224 KB	12.57	12.51	34.37
Den svenska folkbokföringens historia...	434 KB	5.06	5.07	12.78
Genomsnittlig tid (min)		5.55	6.00	15.45

**Tabell 4.2:** Storlek och tid för uppladdning av webbsidor och träning för de olika sentence transformers (ST)

I tabell 4.2 framgår det att det tar längre tid att ladda upp webbsidor till databasen, trots att de inte är lika stora som PDF-filerna. Detta beror på att webbspindeln hanterar alla närliggande webbsidor genom dess hyperlänkar, vilket inte räknas med i storleken. Återigen finns det ingen tydlig parallell mellan storlek och tidsåtgång. Däremot visar resultatet att ST2 tydligt hanterar webbsidor mer effektivt än PDF-filer, vilket kan ses i den markanta skillnaden i tidsåtgång mellan tabell 4.1 och 4.2.

Eftersom syftet med detta system är att kunna extrahera relevant data för användaren, var AI-modellens precision viktig. Därför utfördes tre olika söktester för att få en uppfattning om hur systemet presterar. Tre specifika sökfraser valdes - "redovisning", "Hur gör man ett bokslut?" och "fotboll" - med hänsyn till att flera dokument i databasen berörde dessa områden, samt ett som inte var relevant till något dokument i databasen. Både PDF-filer och webbsidor inkluderades för att maximera testets representativitet. För att etablera en tröskel för relevans användes ett cosine similarity-värde på 0.4 och därefter 0.6 från modellen baserat på söktermen. Det innebär att alla dokument med ett värde över 0.4 respektive 0.6 ansågs som relevanta för sökningen och presenterades för användaren. De sökresultat som ansågs relevanta för söktermen med en tröskel på 0.4 markerades i orange, och de

resultat som uppnådde en tröskel på 0.6 markerades i blått.

Filnamn	ST1	ST2	ST3
Anvisningar bokslut och rapportering	0.4318	0.4104	0.7353
Avdragsrätt, fakturering och redovisning	0.3932	0.3251	0.6031
Beställa i inköps- och fakturasystemet	0.1748	0.3686	0.4990
Bokslutsmaterial	0.1991	0.3686	0.5892
Budget och prognos	0.1917	0.3686	0.5454
Deklarera för föreningar	0.2129	0.2128	0.4924
Den svenska folkbokföringens historia...	0.2298	0.2896	0.4924
Du är medborgare i Storbritannien med...	0.2298	0.3104	0.4924
E-tjänster och blanketter för företag	0.2332	0.2464	0.5026
Fakturera interna och externa kunder	0.2013	0.3686	0.4990
Godkänn leverans och betala din...	0.1292	0.3282	0.5322
Hantera fakturor – vanliga konton vid...	0.2919	0.3698	0.6186
Inkomst av näringsverksamhet	0.2332	0.2332	0.54945
Intäkter	0.3145	0.3947	0.6995
Investeringar och anläggningstillgångar	0.3072	0.4708	0.6773
Kontaktpersoner ekonomistöd	0.2575	0.3686	0.6501
Kontoplan & objektsplan med info...	0.2729	0.3861	0.7076
Kostnader	0.3581	0.3573	0.7029
Moms - Företag och organisationer	0.2945	0.2945	0.4924
Omräkningsstöd indirekta kostnader...	0.3807	0.3334	0.7191
Organisation och uppdrag för inköp	0.2129	0.3736	0.5448
Projektbudget	0.2190	0.3268	0.6878
Projektkalkyl	0.2335	0.3721	0.4860
Rapportera till finansjär	0.2697	0.4660	0.6271
Redovisning & uppföljning av verk...	0.4304	0.4832	0.6850
Redovisningsmodell	0.2333	0.4575	0.6817
Redovisningens syfte & principer	0.3717	0.5362	0.7109
Ränteutgifter	0.2332	0.2332	0.6033
Svar på vanliga frågor - deklARATION...	0.2930	0.2876	0.5027
Systemstöd för ekonomi	0.2733	0.3963	0.7098
Så här gör du inköp	0.1832	0.3844	0.5017
Särskilda skatt- och momsregler	0.1378	0.4243	0.4697
Tidplan ekonomiarbete	0.1878	0.3089	0.6580
Uppgifter och adresser för in- och ut...	0.1615	0.3806	0.6128
Utbildningsfilmer Proceedo	0.2158	0.3232	0.4874
Tid (min)	1:28	1:43	1:24

**Tabell 4.3:** Cosine similarity-värden och uppmätt tid från sökningen “redovisning” för de olika sentence transformers

Resultatet visar att data som är mindre relevanta för sökningen får ett lägre värden jämfört med data som är mer relevanta, vilket leder till ett effektivt stödsystem och eliminerar irrelevant information. Tabellen 4.3 visar även att *KBLab/sentence-bert-swedish-cased* (ST3) [12] presterade bättre än de andra modellerna vid identifiering

av relevant data till söktermen. Eftersom alla dess värden är generellt högre än de andra, visas en indikation på att den är mer effektiv på att identifiera och matcha relevanta dokument för de givna söktermerna. Detta visar på att ST3 har en bättre prestanda när det gäller att förstå och bearbeta de specifika sökfrågorna och dokumentinnehållet i detta sammanhang. *KBLab/sentence-bert-swedish-cased* (ST3) [12] blev därför den modell som slutligen används i projektet.

För de nästa två söktermerna kommer de topp 10 resultaten visas med dess värden. Här kommer endast ST3 användas.

Index	Filnamn	Likhet
1	Inkomst av näringsverksamhet	0.7506
2	Ränteutgifter	0.7506
3	Anvisningar bokslut och rapportering.pdf	0.7093
4	Bokslutsmaterial.pdf	0.6811
5	Redovisningens syfte & principer.pdf	0.6577
6	Redovisningsmodell.pdf	0.6503
7	Avdragsrätt, fakturering och redovisning	0.6491
8	E-tjänster och blanketter för företag	0.6491
9	Moms - Företag och organisationer	0.6491
10	Svar på vanliga frågor - deklaration - privat	0.6400
11	Deklarera för föreningar	0.6039
12	Systemstöd för ekonomi.pdf	0.5954
13	Redovisning & uppföljning av verksamhet.pdf	0.5842
14	Omräkningsstöd indirekta kostnader i funktions...	0.5722
15	Kontoplan & objektsplan med information...	0.5643
16	Hantera fakturor – vanliga konton vid inköp.pdf	0.5521
17	Intäkter.pdf	0.5504
18	Den svenska folkbokföringens historia under...	0.5240
19	Du är medborgare i Storbritannien med uppeh...	0.5240
Tid (min)		1:16

**Tabell 4.4:** Cosine similarity-värden och uppmätt tid från sökningen “Hur gör man ett bokslut?” för ST3

Från sökningen “Hur gör man ett bokslut?” erhöles 23 sökresultat. Eftersom ST3 generellt hade högre värden jämfört med de andra sentence transformers, beslutades att höja tröskeln från 0.4 till 0.6 för att få så relevanta sökresultat som möjligt. Linjen i tabellen 4.4 visar var tröskeln går. Sökresultaten är rangordnade efter relevans, och det framgår att alla har ett högt värde, vilket gör dem relevanta för sökningen.

Den sista söktermen “fotboll” valdes för att illustrera vikten av en tröskel, genom vilken AI-modellen kan extrahera endast relevanta sökresultat. Dokumenten och värdena i tabellen 4.5 visades inte på webbplatsen eftersom de inte var relevanta för söktermen. Istället visades informationsmeddelandet “No results found in the database. Try using a different search phrase.” för användaren. Baserat på dessa resultat är det viktigt att diskutera deras implikationer och vad de innebär för stödsystemets prestanda. En djupare analys av de observerade skillnaderna mellan de olika sen-

tence transformers, samt en utvärdering av deras prestanda och effektivitet i olika sammanhang, presenteras i diskussionskapitlet.

Index	Filnamn	Likhet
1	Kontaktpersoner ekonomistöd.pdf	0.2568
2	Tidplan ekonomiarbete.pdf	0.2447
3	Projektbudget.pdf	0.2051
4	Redovisning & uppföljning av verksamhet.pdf	0.2022
5	Så här gör du inköp.pdf	0.1991
6	Utbildningsfilmer Proceedo.pdf	0.1967
7	Kontoplan & objektsplan med information...	0.1928
8	Organisation och uppdrag för inköp.pdf	0.1896
9	Redovisningens syfte & principer.pdf	0.1883
10	Beställa i inköps- och fakturasystemet Proceedo.pdf	0.1875
11	Svar på vanliga frågor - deklaration - privat	0.1771
12	E-tjänster och blanketter för företag	0.1530
13	Moms - Företag och organisationer	0.1530
14	Ränteutgifter	0.1530
15	Inkomst av näringsverksamhet	0.1530
16	Den svenska folkbokföringens historia under tre...	0.1530
17	Deklarera för föreningar	0.1530
18	Avdragsrätt, fakturering och redovisning	0.1530
19	Du är medborgare i Storbritannien med uppehåll...	0.1530
Tid (min)		1:18

**Tabell 4.5:** Cosine similarity-värden och uppmätt tid från sökningen "fotboll" för ST3.

# 5

## Diskussion

Projektets syfte var att utveckla ett stödsystem för Chalmers ekonomiavdelning, för att underlätta och effektivisera deras arbetsflöde. Systemet möjliggör sökningar i en databas för att bland annat hitta länkar till externa webbsidor, manualer och kontaktpersoner. Söktermen hanteras av en AI-modell som använder bland annat en sentence transformer för att extrahera relevant information från databasen. För att identifiera den mest effektiva modellen för detta ändamål undersöktes flera olika förtränade modeller och jämfördes med andra sökmotorer.

### 5.1 Metoddiskussion

I slutet på projektet identifierades flera svagheter i metoden, och en av de är att det inte utfördes omfattande testning av sambandet mellan det använda relevansmättet och användarupplevd relevans på resultaten. Då syftet med projektet inkluderar att underlätta för anställda på Chalmers bör mer vikt ha lagts på att utföra tester angående hur de upplever systemet gentemot det befintliga, samt hur de upplever relevansen av resultaten. Trots att modellen *KBLab/sentence-bert-swedish-cased* åstadkom en hög relevans enligt cosine similarity-mättet som modellerna utvärderades på, kan det skilja sig till vad användarna, som är kunniga inom datan, upplever vara ett korrekt och relevant sökresultat. Dessa utvärderingsmetoder krävs för att kunna dra en slutsats angående huruvida stödsystemet kan extrahera relevant data effektivt och med precision. Eftersom Chalmers data blev tillgänglig i slutet av projektet, begränsades möjligheten att genomföra omfattande tester av applikationen, vilket är den främsta anledningen till att sådana tester uteblev. Det bör därför vara en hög prioritet att utföra liknande utvärderingar inför vidareutveckling av systemet. Det finns även fler likhetsmått som stödsystemet kan utvärderas på, exempelvis skalärprodukten av vektorerna [23]. Tester för att utvärdera skillnaden på dessa mått och eventuellt optimera stödsystemets precision prioriterades inte heller på grund av tidsbrist.

Ännu en svaghet i metoden kan vara omfattningen av tester på stödsystemets funktioner. Endast tre olika sökord testades, vilka i viss mån gav goda resultat som talade för ett lyckat söksystem. Dock krävs fler tester, som kan innefatta fler möjliga användarfall inom de arbetsuppgifter som en Chalmersanställd möjligen genomför. Med fler tester så hade en tydligare slutsats kunnat dras, angående hur robust systemet är. Även i detta fall så var tidsbegränsningen faktorn till varför så få tester genomfördes. Databasen hade även kunnat utökas mer, möjligen med mer varierad och

mångfaldig data, för att testerna skulle bli mer trovärdiga. En utökning av detta område hade gett ytterligare stöd till att kunna dra pålitliga slutsatser.

Beslutet att använda förtränade modeller gjorde att tid kunde sparas, då modellen hade en grundförståelse av språkliga mönster. Dock är det inte en garanti på att modellen ska vara utformad för just den data som den tränades med under projektets gång. Detta kan leda till svårigheter med att förstå specifika facktermer inom ekonomi och administration, som kan vara avgörande för modellens korrekthet. Genom att ytterligare träna dessa modeller på relevanta termer utöver de som förekom i Chalmers data, kan resultaten förbättras. Antalet förtränade modeller på svenska språket var begränsat, och tillsammans med tidsbegränsningen var det inte möjligt att utöka testningen till fler modeller. Alternativt kunde en modell ha byggts upp från grunden och endast tränats på datan från Chalmers samt relevanta sidor som Skatteverket och Bolagsverket, men faktumet att Chalmers data dröjde satte hinder för att kunna utföra detta. Dessa relevanta sidor kan även samlas in från Chalmers anställda på de externa webbsidor som de använder i deras arbetsuppgifter.

## 5.2 Val av Sentence Transformer

Valet av sentence transformer är avgörande för att säkerställa att stödsystemet kan tolka användarens sökningar korrekt och effektivt. Efter att ha testat tre modeller - *all-MiniLM-L6-v2* (ST1)[10], *all-mpnet-base-v2* (ST2)[11] och *KBLab/sentence-bert-swedish-cased* (ST3)[12] - visade resultaten att ST3 presterade bäst. Detta kan motiveras av flera faktorer, som modellens dimensionsantal och språkträning.

Modellen *all-MiniLM-L6-v2* kartlägger texter till ett 384-dimensionellt vektorrum [10]. Detta är det lägsta antalet av de tre modellerna som utvärderas, vilket gör den till den mest effektiva när det gäller beräkningstid och resursanvändning, vilket visas vid uppladdning av PDF-filer i tabell 4.1. Den lägre dimensionen innebär att modellen inte behöver bearbeta lika många dimensioner, vilket minskar beräkningstiden, och samtidigt precisionen.

De två andra modellerna, *all-mpnet-base-v2* och *KBLab/sentence-bert-swedish-cased*, använder båda ett 768-dimensionellt vektorrum. Dessa modeller är något långsammare än ST1 när det kommer till beräkningstid, men de ger bättre representationer av semantiska relationer på grund av det högre dimensionsantalet [11, 12, 24]. Detta kan ses i tabell 4.3 där det tydligt syns en ökning. Emellertid är ST2 inte lika bra på att identifiera semantiska likheter mellan sökfraser och hemsidor som mellan sökfraser och PDF-filer. Detta kan bero på att språket som används på webbsidorna skiljer sig från träningsdatans språk, då *all-mpnet-base-v2* (ST2) [11] endast är tränad på engelsk data.

En av de främsta anledningarna till att *KBLab/sentence-bert-swedish-cased* presterade bäst, och därmed valdes att användas till detta projekt, är att den är tränad på både svenska och engelska. Detta gör den optimal för ett stödsystem i Sverige där majoriteten av databasen innehåller svensk data, men där engelska också kan

förekomma. Den tvåspråkiga träningen innebär att modellen kan hantera text på båda språken effektivt. Dessutom har *KBLab/sentence-bert-swedish-cased* tränats med *all-mpnet-base-v2* som lärarmodell, vilket innebär att den har tillgång till den språkliga förståelsen från ST2, men med ytterligare kunskap från svenska språket [12].

### 5.3 Jämförelse med andra sökmotorer

För att kunna utvärdera applikationen så jämfördes den med en liknande sökmotor, AISeek. AISeek är utformad för att fungera som t.ex. ChatGPT, men utan att samla in användardata. AISeek-appen använder sig av ChatGPT, men kan även anpassas till andra AI-modeller. ChatGPT är byggd med AI-algoritmer och tränad på stora mängder data för att kunna svara på användarens frågor. I appen kan användaren ställa diverse frågor, begränsade till modellens informationskälla, vilket är all tillgänglig information på internet. Dessutom har appen förmågan att söka genom internet i realtid, vilket gör att den kan bistå med aktuell information [25]. Jämförelsen gjordes genom att studera AISeek och dess funktioner, och sedan jämföra dessa med det utvecklade stödsystemet. Eftersom AISeek är en betaltjänst, baserades jämförelsen på tillgänglig information om appen, och inte på praktisk användning.

Skillnaden mellan applikationen i detta projekt och AISeek är att användaren själv måste lägga in specifika dokument och webbsidor som AI-modellen tränas på. Detta innebär att AISeek kan ha större kapacitet att bearbeta sökfraser och generera resultat, eftersom den har tillgång till ett bredare utbud av information. Dock är detta projekt riktat mot en specifik användargrupp, Chalmers ekonomiavdelning, vilket innebär att den information som söks ofta inte är offentlig och därför inte tillgänglig för AISeek. För att få tillgång till exempelvis kontaktuppgifter till personal inom Chalmers kan det krävas inloggning av auktoriserad personal. Med detta i åtanke är AISeek inte lämplig för detta specifika syfte som detta stödsystem är utformad för, men kan vara lämplig för andra avsikter eller integreras med webbsökningen av projektet. Detta skulle då ge ett kraftfullt stödsystem med internet som sökbank, där användaren själv kan komplettera med egna dokument.

En annan uppmärksam skillnad är att svaret på en fråga från AISeek är ett formulerat svar, som kan vara mer nyanserat och komplext, medan svaren hos detta stödsystem är enklare, där användaren själv får läsa igenom den webbsida eller PDF-fil som visas i resultatet. Ännu en aspekt är den med säkerheten och anonymiteten. AISeek har ett väl utvecklat system för att hålla användare anonyma och inte samla in användardata, medan detta inte haft ett stort fokus i utvecklingen av detta stödsystem. Trots detta implementerades ingen inloggning i applikationen, vilket säkerställer att användardata inte samlas in, och därmed inte riskeras att läcka.

## 5.4 Förbättringsområden och vidareutveckling

För att ett stödsystem ska effektivisera organisationens arbetsflöde måste det vara både snabbt och precist. Därför är ett potentiellt förbättringsområde testning av olika typer av AI-modeller och Sentence Transformers. Genom att jämföra dessa modeller kan den mest effektiva och precisa modellen hittas för användning till stödsystemet. Även att kontrollera flera olika mått för att hitta den optimala modellen hade kunnat förbättra resultatet. En annan metod att överväga är användningen av K-Nearest Neighbors (KNN) algoritmen för semantisk sökning. KNN kan hjälpa till att identifiera och rangordna dokument baserat på deras semantiska likhet med sökfrasen, vilket kan ytterligare förbättra sökprecisionen [14].

Under projektets gång har vissa buggar uppstått som kan påverka användarupplevelsen, men som inte påverkar programmets funktionalitet. På grund av tidsbrist har dessa buggar inte åtgärdats. Ett exempel är begränsningen att användaren endast kan ladda upp PDF-filer upp till totalt 10 MB per uppladdning. Detta kan vara ett problem när stora mängder data krävs, särskilt om en enskild PDF-fil överstiger gränsen. Den största filen mottagen från Chalmers var 5,8 MB, vilket gjorde att denna bugg inte prioriterades.

Utseendet på hemsidan kan också förbättras ytterligare för att öka användarvänligheten. Genom att utveckla designen kan användarupplevelsen optimeras och därmed kan organisationens arbetsflöde effektiviseras ytterligare. Webbsidan kan även vidareutvecklas för att se historik på söktermer och även hantera stavfel i söktermen. Applikationen kan även utvecklas så att webbsidan kan nås överallt, inte bara lokalt. Ännu ett område som kan utvecklas vidare är möjligheten att kunna logga in och ha enskilda användare. På så sätt kan användaren ha sin privata databas. Vidare kan möjligheten för användare att även ta bort innehåll från databasen utforskas och möjligen utvecklas.

## 5.5 Etik och miljö

I utveckling och implementering av detta projekt är det viktigt att diskutera etiska och miljöaspekter. De aspekter som beaktats inom etiken är utifrån användarens säkerhet, integritet och upplevelse, resultaten, och huruvida den kan ersätta jobb. De aspekter som beaktats inom miljö är bland annat energiåtgång och koldioxidutsläpp.

### 5.5.1 Etik

En del av det som är viktigt ur ett etiskt perspektiv är att säkerställa användares data och privata information. Eftersom möjligheten om att ladda upp privata PDF-filer finns, så kan det uppstå en risk att dessa inte förvaras på ett sätt som respekterar användarens integritet. Det bör säkerställas att åtkomsten till PDF-filerna från databasen är helt begränsad, och applikationen bör ha skydd mot dataintrång såsom SQL-injektion. Genom att använda Django så har applikationen ett inbyggt skydd mot dataintrång och datan i databasen förvaras på ett tryggt sätt. Vidare

så bör användaren bli informerad om vilken data som samlas in, och hur den används. Eftersom applikationen inte är baserad på inloggade användare, kommer inte informationen som laddas upp att kopplas till någon specifik person, däremot kommer den att vara kvar i den lokala databasen tills den tas bort därifrån manuellt. Dessutom sparas ingen data om varken IP-adress eller position av den enhet som applikationen används på, vilket förstärker garantin om att användare förblir anonyma. Det kan komma att bli en säkerhetsbrist då känslig information kan vara lagrad i databasen vid implementering av användare. I det fall är det viktigt att implementera ytterligare säkerhetsåtgärder, såsom kryptering av lagrad data, och regelbundna säkerhetsgranskningar, för att minimera risken för dataläckor och obehörig åtkomst.

Ytterligare en aspekt angående sökresultaten är att dessa kan vara partiska beroende på hur träningsdatan ser ut. Om träningsdatan är likformig och inte nyanserad så kan sökresultaten komma att vara partiska. Det är emellertid applikationens funktion att användaren själv väljer ut vad den vill lägga in i databasen, och därmed träna AI-modellen med. Det kan även medföra en risk för att sökresultaten visar på falsk information, beroende på vad användaren väljer att lägga till. Här har användaren stor frihet och därmed stort ansvar att förse applikationen med korrekt data. Det kan vara fördelaktigt att säkerställa att användaren är medveten om att resultaten endast kommer spegla den data som manuellt lagts in. I denna version av programmet har det inte varit en prioritet, men kan komma att läggas in i senare versioner ifall det anses avgöra upplevelsen för användaren.

För att inte bryta mot etiska normer bör applikationen vara lättillgänglig för alla och inte diskriminera mot någon. Detta kan göras genom att se till att gränssnittet på hemsidan är användarvänligt och inte främjar någon slags nedsättande behandling. Det kan även göras genom att se till att applikationen ska vara gratis, vilket denna version är, eller ha ett överkomligt pris. I nuläget finns inget på hemsidan som kan diskriminera, då den endast innehåller det nödvändiga för att ladda upp data och söka på sökfraser, samt en sektion för resultaten. Dock bör ändå möjligheten att rapportera och anmäla problem eller bekymrade upptäckter finnas för att säkerställa att applikationen är transparent och att användaren värdesätts genom hela dess upplevelse. Vidare bör applikation inte ersätta jobb om den används på en arbetsplats. Eftersom det i dagsläget inte finns ett specifikt jobb kopplat till att söka information på Chalmers intranät finns det ingen risk för att eliminera en anställning på grund av denna applikation vid användning där. Applikationens syfte är endast att vara ett stöd- och navigeringssystem för de anställda för att hitta information enklare.

### 5.5.2 Miljö

En aspekt ur miljösyn är angående energiåtgång och koldioxidutsläpp. Driften för att hantera AI-modeller och hantera sökfrågor samt lagra användardata kan kräva stora mängder energi, vilket visats under resultatkapitlet. I testerna för att köra de olika sentence-transformers uppmättes den längsta tiden till 34.37 minuter. Beroende på hur mycket data som användaren vill ladda upp så kan det resultera i en stor

mängd energiåtgång. För att inte bidra till ökat koldioxidavtryck är det viktigt att koden bakom applikationen är så optimerad och effektiv som möjligt, så att datorn inte måste arbeta i onödan. Den ska kräva så lite beräkningskraft som möjligt för att samtidigt kunna uppehålla sin prestanda och funktionalitet, samt användarvänlighet. Detta har försökts att implementeras under projektets gång, men vid vidare utveckling så kan koden revideras och skalas av ytterligare. Applikationen bör inte skada den hårdvara den körs på, för att inte öka det elektroniska avfallet. Det kan vara aktuellt att utföra en livscykelanalys för att kunna identifiera miljöpåverkan för applikationen, från utveckling till drift och avveckling. Denna analys kan hjälpa att identifiera negativa miljöeffekter, som i sin tur kan arbetas för att minskas.

# 6

## Slutsats

Detta examensarbete har tagit fram ett AI-drivet stödsystem för att förenkla navigeringen på Chalmers intranät. Genom att implementera AI-modeller har systemet visat potential till att extrahera relevant information från en databas, baserat på användarens söktermer. Denna potential har demonstreras genom cosine similarity-värden som beräknas från diverse tester för att undersöka systemets prestanda. För att styrka dessa resultat krävs dock omfattande tester med den avsedda användargruppen för att säkerställa vad som betraktas som relevant information. En slutsats gällande sentence transformers är att *KBLab/sentence-bert-swedish-cased* var den mest effektiva modellen av de som testades under projektets gång, tack vare dess förmåga att hantera både svenska och engelska. Ytterligare utvärderingar och tester är nödvändiga för att bekräfta dessa slutsatser och optimera systemet för praktisk användning.

Framtida versioner av systemet kan utöka dess kapacitet att lagra och hantera diverse webbsidor och dokument, samt förbättra AI-modellerna för att ytterligare optimera användarupplevelsen. Ytterligare säkerhetsåtgärder bör övervägas för att skydda känslig data, och användargränssnittet kan förbättras för att bli ännu mer intuitivt och användarvänligt. Dessutom kan en livscykelanalys genomföras för att minimera miljöpåverkan. Sammanfattningsvis har detta projekt lagt grunden för ett effektivt och användarvänligt stödsystem som har stor potential för framtida utveckling och anpassning till användarnas behov.



# Litteraturförteckning

- [1] Canva AI, “AI-generated image of a computer,” Canva, 2024. [Online]. Available: <https://www.canva.com>. (accessed Jun. 02, 2024).
- [2] R. Roy, “Best Python libraries for Machine Learning,” GeeksforGeeks, Jan. 18, 2019. [Online]. Available: <https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/>. (accessed Jun. 02, 2024).
- [3] “8 Reasons Why Python is Good for Artificial Intelligence and Machine Learning,” Django Stars Blog, Mar. 11, 2019. [Online]. Available: <https://djangostars.com/blog/why-python-is-good-for-artificial-intelligence-and-machine-learning/>. (accessed Jun. 02, 2024).
- [4] Django Software Foundation, “The Web framework for perfectionists with deadlines,” *Django*. [Online]. Available: <https://www.djangoproject.com/>. (accessed May 21, 2024).
- [5] K. R. Chowdhary, *Fundamentals Of Artificial Intelligence*. Springer, 2020.
- [6] “Feature Extraction Explained,” se.mathworks.com. [Online]. Available: <https://se.mathworks.com/discovery/feature-extraction.html>. (accessed Jun. 02, 2024).
- [7] “GPT-SW3 | AI Sweden,” [Online]. Available: <https://www.ai.se/en/project/gpt-sw3>. (accessed Jun. 02, 2024).
- [8] “AI-Sweden-Models/gpt-sw3-126m-instruct · Hugging Face,” huggingface.co, Jun. 30, 2023. [Online]. Available: <https://huggingface.co/AI-Sweden-Models/gpt-sw3-126m-instruct>. (accessed Jun. 02, 2024).
- [9] “SentenceTransformers Documentation — Sentence Transformers documentation,” sbert.net. [Online]. Available: <https://sbert.net>. (accessed Jun. 02, 2024).
- [10] “sentence-transformers/all-MiniLM-L6-v2,” huggingface.co. [Online]. Available: <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>. (accessed Mar. 12, 2024).
- [11] “sentence-transformers/all-mpnet-base-v2,” huggingface.co. [Online]. Available: <https://huggingface.co/sentence-transformers/all-mpnet-base-v2>. (accessed May 21, 2024).
- [12] “KBLab/sentence-bert-swedish-cased,” huggingface.co, Dec. 18, 2023. [Online]. Available: <https://huggingface.co/KBLab/sentence-bert-swedish-cased>. (accessed May 21, 2024).
- [13] Hugging Face, “Transformers,” *PyPI*. [Online]. Available: <https://pypi.org/project/transformers/>. (accessed May 21, 2024).

- [14] S. Shirol, A. Kulkarni and R. Agarwal, “Semantic Search for Sustainable Platforms Using Transformers,” 2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), Hyderabad, India, 2023, pp. 112-118, doi: 10.1109/ICETCI58599.2023.10331079. (accessed May 21, 2024).
- [15] “Using Sentence Transformers at Hugging Face,” huggingface.co. [Online]. Available: <https://huggingface.co/docs/hub/sentence-transformers>. (accessed Jun. 02, 2024).
- [16] K. Park, J. S. Hong, and W. Kim, “A Methodology Combining Cosine Similarity with Classifier for Text Classification,” Applied Artificial Intelligence, vol. 34, no. 5, pp. 396–411, Feb. 2020, doi: <https://doi.org/10.1080/08839514.2020.1723868>. (accessed Jun. 13, 2024).
- [17] F. vom Lehn, “Understanding Vector Similarity for Machine Learning,” Advanced Deep Learning, May 31, 2024. [Online]. Available: <https://medium.com/advanced-deep-learning/understanding-vector-similarity-b9c10f7506de> (accessed Jun. 16, 2024).
- [18] L. Yu et al., “Summary of web crawler technology research,” *Journal of Physics: Conference Series*, vol. 1449, 2nd International Symposium on Power Electronics and Control Engineering (ISPECE 2019), Tianjin, China, 22-24 Nov. 2019, p. 012036, 2020. Available: <https://iopscience.iop.org/article/10.1088/1742-6596/1449/1/012036/meta>. (accessed Jun. 02, 2024).
- [19] Scrapy, “Overview,” *Scrapy Documentation*. [Online]. Available: <https://docs.scrapy.org/en/latest/intro/overview.html>. (accessed May 21, 2024).
- [20] “Skolämnen i Sverige,” Wikipedia, Aug. 06, 2023. [Online]. Available: [https://sv.wikipedia.org/wiki/Skol%C3%A4mnen\\_i\\_Sverige](https://sv.wikipedia.org/wiki/Skol%C3%A4mnen_i_Sverige). (accessed Jun. 03, 2024).
- [21] S. skatteverket.se, “Startsida,” skatteverket.se. [Online]. Available: <https://skatteverket.se>. (accessed Jun. 03, 2024).
- [22] “Bolagsverket,” Bolagsverket. [Online]. Available: <https://bolagsverket.se>. (accessed Jun. 03, 2024).
- [23] “Semantic Textual Similarity — Sentence Transformers documentation,” www.sbert.net. [Online]. Available: [https://www.sbert.net/docs/sentence\\_transformer/usage/semantic\\_textual\\_similarity.html](https://www.sbert.net/docs/sentence_transformer/usage/semantic_textual_similarity.html) (accessed Jun. 16, 2024).
- [24] “The Basics Of Vector Space Models For SEO,” marketbrew.ai. [Online]. Available: <https://marketbrew.ai/a/vector-space-models-seo>. (accessed Jun. 03, 2024).
- [25] “AI Seek – A revolutionary search application that harnesses the power of AI,” www.aiseek.ai. [Online]. Available: <https://www.aiseek.ai>. (accessed Jun. 03, 2024).

# A

## Appendix 1

<b>Filnamn</b>	<b>Filstorlek</b>	<b>ST1 (s)</b>	<b>ST2 (s)</b>	<b>ST3 (s)</b>
Redovisningens syfte & principer	278 KB	29.51	40.95	42.24
Särskilda skatt- och momsregler	481 KB	31.86	45.63	47.69
Hantera fakturor – vanliga konto...	511 KB	28.31	47.26	51.82
Tidplan ekonomiarbete	844 KB	31.14	40.09	42.49
Kontoplan & objektsplan med info...	1,5 MB	32.84	51.35	51.36
Så här gör du inköp	2 MB	30.92	47.40	48.85
Kontaktpersoner ekonomistöd	5,3 MB	24.43	43.92	45.66
Projektkalkyl	5,8 MB	36.49	49.71	49.98
Genomsnittlig tid (s)		30.69	45.79	47.51

**Tabell A.1:** Filstorlek och tid för uppladdning av PDF-filer och träning för de olika sentence transformers (ST)

<b>Titel på hemsida</b>	<b>Storlek</b>	<b>ST1 (min)</b>	<b>ST2 (min)</b>	<b>ST3 (min)</b>
Deklarera för föreningar	113 KB	3.68	3.69	9.18
E-tjänster och blanketter för företag	118 KB	3.91	3.99	10.03
Avdragsrätt, fakturering och redovisning	125 KB	5.55	5.41	13.42
Moms - Företag och organisationer	128 KB	6.23	4.33	10.73
Inkomst av näringsverksamhet	140 KB	6.98	6.95	17.31
Du är medborgare i Storbritannien med...	162 KB	5.83	5.90	14.90
Ränteutgifter	176 KB	0.14	6.12	16.36
Svar på vanliga frågor - deklaration...	224 KB	12.57	12.51	34.37
Den svenska folkbokföringens historia...	434 KB	5.06	5.07	12.78
Genomsnittlig tid (min)		5.55	6.00	15.45

**Tabell A.2:** Storlek och tid för uppladdning av webbsidor och träning för de olika sentence transformers (ST)

Filnamn	ST1	ST2	ST3
Anvisningar bokslut och rapportering	0.4318	0.4104	0.7353
Avdragsrätt, fakturering och redovisning	0.3932	0.3251	0.6031
Beställa i inköps- och fakturasystemet	0.1748	0.3686	0.4990
Bokslutsmaterial	0.1991	0.3686	0.5892
Budget och prognos	0.1917	0.3686	0.5454
Deklarera för föreningar	0.2129	0.2128	0.4924
Den svenska folkbokföringens historia...	0.2298	0.2896	0.4924
Du är medborgare i Storbritannien med...	0.2298	0.3104	0.4924
E-tjänster och blanketter för företag	0.2332	0.2464	0.5026
Fakturera interna och externa kunder	0.2013	0.3686	0.4990
Godkänn leverans och betala din...	0.1292	0.3282	0.5322
Hantera fakturor – vanliga konton vid...	0.2919	0.3698	0.6186
Inkomst av näringsverksamhet	0.2332	0.2332	0.54945
Intäkter	0.3145	0.3947	0.6995
Investeringar och anläggningstillgångar	0.3072	0.4708	0.6773
Kontaktpersoner ekonomistöd	0.2575	0.3686	0.6501
Kontoplan & objektsplan med info...	0.2729	0.3861	0.7076
Kostnader	0.3581	0.3573	0.7029
Moms - Företag och organisationer	0.2945	0.2945	0.4924
Omräkningsstöd indirekta kostnader...	0.3807	0.3334	0.7191
Organisation och uppdrag för inköp	0.2129	0.3736	0.5448
Projektbudget	0.2190	0.3268	0.6878
Projektkalkyl	0.2335	0.3721	0.4860
Rapportera till finansiär	0.2697	0.4660	0.6271
Redovisning & uppföljning av verk...	0.4304	0.4832	0.6850
Redovisningsmodell	0.2333	0.4575	0.6817
Redovisningens syfte & principer	0.3717	0.5362	0.7109
Ränteutgifter	0.2332	0.2332	0.6033
Svar på vanliga frågor - deklaration...	0.2930	0.2876	0.5027
Systemstöd för ekonomi	0.2733	0.3963	0.7098
Så här gör du inköp	0.1832	0.3844	0.5017
Särskilda skatt- och momsregler	0.1378	0.4243	0.4697
Tidplan ekonomiarbete	0.1878	0.3089	0.6580
Uppgifter och adresser för in- och ut...	0.1615	0.3806	0.6128
Utbildningsfilmer Proceedo	0.2158	0.3232	0.4874
Tid (min)	1:28	1:43	1:24

**Tabell A.3:** Cosine similarity-värden och uppmätt tid från sökningen “redovisning” för de olika sentence transformers

Index	Filnamn	Likhet
1	Inkomst av näringsverksamhet	0.7506
2	Ränteutgifter	0.7506
3	Anvisningar bokslut och rapportering.pdf	0.7093
4	Bokslutsmaterial.pdf	0.6811
5	Redovisningens syfte & principer.pdf	0.6577
6	Redovisningsmodell.pdf	0.6503
7	Avdragsrätt, fakturering och redovisning	0.6491
8	E-tjänster och blanketter för företag	0.6491
9	Moms - Företag och organisationer	0.6491
10	Svar på vanliga frågor - deklaration...	0.6400
11	Deklarera för föreningar	0.6039
12	Systemstöd för ekonomi.pdf	0.5954
13	Redovisning & uppföljning av verksamhet.pdf	0.5842
14	Omräkningsstöd indirekta kostnader i funktions...	0.5722
15	Kontoplan & objektsplan med information...	0.5643
16	Hantera fakturor – vanliga konton vid inköp.pdf	0.5521
17	Intäkter.pdf	0.5504
18	Den svenska folkbokföringens historia...	0.5240
19	Du är medborgare i Storbritannien med...	0.5240
Tid (min)		1:16

**Tabell A.4:** Cosine similarity-värden och uppmätt tid från sökningen “Hur gör man ett bokslut?” för ST3

<b>Index</b>	<b>Filnamn</b>	<b>Likhet</b>
1	Kontaktpersoner ekonomistöd.pdf	0.2568
2	Tidplan ekonomiarbete.pdf	0.2447
3	Projektbudget.pdf	0.2051
4	Redovisning & uppföljning av verksamhet.pdf	0.2022
5	Så här gör du inköp.pdf	0.1991
6	Utbildningsfilmer Proceedo.pdf	0.1967
7	Kontoplan & objektsplan med information...	0.1928
8	Organisation och uppdrag för inköp.pdf	0.1896
9	Redovisningens syfte & principer.pdf	0.1883
10	Beställa i inköps- och fakturasystemet Proceedo.pdf	0.1875
11	Svar på vanliga frågor - deklaration...	0.1771
12	E-tjänster och blanketter för företag	0.1530
13	Moms - Företag och organisationer	0.1530
14	Ränteutgifter	0.1530
15	Inkomst av näringsverksamhet	0.1530
16	Den svenska folkbokföringens historia...	0.1530
17	Deklarera för föreningar	0.1530
18	Avdragsrätt, fakturering och redovisning	0.1530
19	Du är medborgare i Storbritannien med...	0.1530
Tid (min)		1:18

**Tabell A.5:** Cosine similarity-värden och uppmätt tid från sökningen “fotboll” för ST3.

INSTITUTIONEN FÖR DATA- OCH INFORMATIONSTEKNIK  
CHALMERS TEKNISKA HÖGSKOLA  
Göteborg, Sverige  
[www.chalmers.se](http://www.chalmers.se)



GÖTEBORGS  
UNIVERSITET

---



**CHALMERS**