

Evaluering af principper indenfor semantisk  
zoomning

Menneske-datamaskine interaktion

**Thomas René Sidor - thomas@mbyte.dk**

Vejleder: Kasper Hornbæk

*Datalogisk Institut Københavns Universitet, Forår 2006*

## Indhold

<b>1</b>	<b>Begrebet semantisk zoomning</b>	<b>8</b>
1.1	Anvendelsesmuligheder . . . . .	9
1.2	Generelle problemstillinger indenfor semantisk zoomning . . . . .	9
<b>2</b>	<b>Problemformulering</b>	<b>10</b>
2.1	Afgrænsning . . . . .	10
2.2	Motivation . . . . .	11
<b>3</b>	<b>Identificering af opgaver</b>	<b>12</b>
3.1	Opgavedefinitionen . . . . .	12
3.1.1	Overordnet trend . . . . .	13
3.1.2	Opdeling i delperioder . . . . .	13
3.2	Antagelser om datatypen . . . . .	14
<b>4</b>	<b>Metoder til semantisk zoomning på statistiske dataark</b>	<b>15</b>
4.1	Generelt . . . . .	15
4.2	Gruppering . . . . .	16
4.3	Filtrering . . . . .	16
4.4	Sortering . . . . .	17
4.5	Generelt om sammenfletning af celler . . . . .	17
4.5.1	Vertikal sammenfletning . . . . .	17
4.5.2	Horisontal sammenfletning . . . . .	18
4.6	Repræsentationer på celleniveau . . . . .	19
<b>5</b>	<b>Beskrivelse af løsningsforslag</b>	<b>21</b>
5.1	Overordnede metoder . . . . .	21
5.1.1	Rækkeoperationer . . . . .	21
5.1.2	Kolonneoperationer . . . . .	22
5.1.3	Celleoperationer . . . . .	23
5.2	Overgange mellem repræsentationsformer . . . . .	23
5.2.1	Interessegrad for celler . . . . .	24
5.2.2	Interessegrad for rækker . . . . .	25
<b>6</b>	<b>Implementering af udvalgte metoder</b>	<b>26</b>
6.1	Løsning med geometrisk zoomning . . . . .	26
6.2	Løsning med semantisk zoomning . . . . .	27
6.3	Regnemaskine . . . . .	28
<b>7</b>	<b>Evaluering af implementerede metoder</b>	<b>28</b>
7.1	Kvantificering af metodernes brugbarhed . . . . .	28
7.2	Udvalgt sammenligningsgrundlag . . . . .	29
7.3	Evalueringsform . . . . .	29
7.3.1	Introduktion . . . . .	29

7.3.2	Opgaver . . . . .	31
7.3.3	Spørgeskema . . . . .	33
7.3.4	Efter forsøget . . . . .	34
7.3.5	Udstyr . . . . .	34
7.3.6	Forsøgspersoner . . . . .	34
<b>8</b>	<b>Resultat af evaluering</b>	<b>35</b>
8.1	Dataanalyse . . . . .	35
8.1.1	Effektivitet . . . . .	35
8.1.2	Fejlrate . . . . .	36
8.1.3	Subjektiv undersøgelse . . . . .	36
8.1.4	Præference . . . . .	38
8.1.5	Erfaringsniveau . . . . .	38
8.1.6	Angivne fordele . . . . .	38
8.2	Resultat . . . . .	38
8.2.1	Videre undersøgelser . . . . .	39
<b>A</b>	<b>Bilag</b>	<b>41</b>
A.1	Skærmbilleder . . . . .	41
A.2	Skærmbilleder fra grænseflade med semantisk zoomning . . . . .	42
A.3	Dataark til brugertests . . . . .	45
A.4	Eksempel på svarark til spørgsmål . . . . .	49
A.5	Spørgeskema . . . . .	50
A.6	Datagrundlag for effektivitet . . . . .	52
A.7	Datagrundlag til fejlrate . . . . .	53
A.8	Datagrundlag for subjektiv analyse . . . . .	54

## Figurer

1	Dataserie med angivelse af overordnet trend . . . . .	13
2	Dataserie opdelt i delperioder . . . . .	14
3	Referenceløsning . . . . .	15
4	Sortering af rækker . . . . .	17
5	Eksempel på vertikal sammenfletning af celler . . . . .	18
6	Eksempel på horisontal sammenfletning af celler . . . . .	19
7	Cellerepræsentation, referenceløsning . . . . .	19
8	Cellerepræsentation, absolut ændring . . . . .	19
9	Cellerepræsentation, relativ ændring . . . . .	20
10	Cellerepræsentation, enhedsskift . . . . .	20
11	Cellerepræsentation, interval . . . . .	20
12	Cellerepræsentation, informationer . . . . .	21
13	sammenfletning af kolonner. Øverst ses dataarket forinden søjlen sammenflettes og nederst efter kollapset. . . . .	22
14	Interessegraden udtrykt ved forskellige zoomniveauer samt græn- seværdi angivet ved stiplede linje . . . . .	25
15	Skærbillede af grænseflade med semantisk zoomning . . . . .	27
16	Skærbillede af dataark på laveste zoomniveau . . . . .	28
17	Referenceløsningens grænseflade . . . . .	41
18	Regnemaskinens grænseflade . . . . .	42
19	Skaleringseksempel 1, Semantisk zoomning . . . . .	42
20	Skaleringseksempel 2, Semantisk zoomning . . . . .	43
21	Skaleringseksempel 3, Semantisk zoomning . . . . .	43
22	Skaleringseksempel 4, Semantisk zoomning . . . . .	44
23	Eksempelside af opgaveark med svarplads . . . . .	49
24	Spørgeskema til subjektiv undersøgelse, side 1 og 2 . . . . .	50
25	Spørgeskema til subjektiv undersøgelse, side 3 . . . . .	51

## Tabeller

1	Rækkefølge af grænseflade og opgavesæt ved forsøg . . . . .	31
2	Gennemsnitlig opgavegennemførelsestid, standardafvigelse her- fra og $p$ for de to grænseflader . . . . .	35
3	Gennemsnitlig fejlrate, standardafvigelse herfra og $p$ for de to grænseflader . . . . .	36
4	Gennemsnitlig tilfredshedsvurdering (standardafvigelse) af de to grænseflader . . . . .	37
5	Subjektiv vurdering af grænsefladerne, $p$ -værdier . . . . .	37
6	Grupper af fordele ved grænseflade med semantisk zoomning .	38
7	Gennemsnitlig opgavegennemførelsestid, standardafvigelse her- fra og $p$ for de to grænseflader fordelt på de enkelte opgaver .	52

8	Konstaterede fejl ved opgaveløsning . . . . .	53
9	Fejlrate for de to grænseflader fordelt på de enkelte opgaver .	53
10	Datagrundlag for subjektiv analyse . . . . .	54

## Indledning

Princippet om semantisk zoomning er et relativt uudforsket koncept. Om end det siden introduktionen i 1993 er blevet forsøgt benyttet i en række forskellige sammenhænge, er der endnu mange områder, som endnu ikke har fået lejlighed til at afprøve konceptet.

Denne opgave søger at udforske dette koncept i en endnu ikke før afprøvet situation - nemlig benyttet på statistiske datasæt præsenteret i regnearksform.

Der er ved identificering af brugernes tænkelige opgaver, kraftigt baseret på erfaringer gjort i kurset Samfundsbeskrivelse ved Økonomisk Institut på Københavns Universitet, fundet frem til at det er indledende arbejde med trendanalyse og delperiodeopdeling som søges støttet. Ved implementering af et løsningsforslag til en grænseflade som gør brug af semantisk zoomning - med diverse former for række- og kolonnesammenfletninger, intervalberegninger og andet - er princippet om semantisk zoomning blevet søgt sammenlignet med dens forfader - almindelig geometrisk zoomning.

Ved evaluering af to grænseflader - den udviklede og en referenceløsning - er der da søgt belyst nogle målbare egenskaber som giver et indtryk af hvordan dette koncept fungerer i virkeligheden. Disse egenskaber er således opgavegennemførelsestid, fejlrate og subjektiv tilfredshed. Ved evalueringen, baseret på brugerforsøg med syv brugere, fandtes det at der umiddelbart endnu ikke kan siges noget klart om forskellen på de to løsninger. Dataene peger dog imod at opgaver med fokus på delperiodeopdeling klares væsentligt bedre i grænsefladen med semantisk zoomning - med gennemførelsestider 31 % hurtigere og en fejlrate som er markant lavere. Dette er endvidere noget som undersøttes af brugernes subjektive vurdering hvor forsøgspersonerne fandt grænsefladen mere behjælpelig i forbindelse med at skabe et overblik over de data der fremstilles for dem.

Det må dog siges at der er behov for yderligere undersøgelser for at kaste lys på nogle af de uklare resultater som fremstilles i denne opgave.

**Problemformulering** Det er valgt at ændre problemformuleringen i denne opgave fra den oprindelige i synopsisen. Således er den ændret fra værende "I hvilket omfang støtter udvalgte principper indenfor semantisk zoomning brugere i deres opgaver?" til værende følgende

*"I hvilken grad støtter princippet om semantisk zoomning brugere i deres arbejde med indledende trendanalyse og delperiodeopdeling af større statistiske dataark repræsenteret i regnearksform."*

Dette er valgt for at understrege hvilken brugssituation der er tale om samt hvilke opgaver der søges støttet. Således findes den nye problemformulering bedre i forhold til indholdet i denne opgave samtidig med at der kan svares mere præcist herpå.

## Læsevejledning

Ved læsning af denne rapport vil du mere eller mindre linært blive ført igennem den proces der leder mod resultatet af undersøgelserne. Således vil der kort blive redegjort for begrebet semantisk zoomning efterfulgt af opgavens problemformulering, afgrænsning og motivation. Herefter søges det at identificere brugernes opgaver med dataene som ønskes støttet, hvorefter generelle metoder til semantisk zoomning på sådanne data diskuteres. Den implementerede løsning samt referenceløsningen vil da blive beskrevet hvorefter der redegøres for de evalueringmetoder som er taget i brug - dette endeligt afsluttet med en dataanalyse samt en opsummering af resultaterne heri samt forslag til videre undersøgelser.

## 1 Begrebet semantisk zoomning

For at forklare begrebet, lad mig da først komme med et velkendt eksempel - almindelige by- og landekort. Et kort i forholdet 1:1 ville ikke give megen mening - derfor har vi med kartografien introduceret en række bredt accepterede repræsentationer for de ting vi møder i vores hverdag. Alt fra træer, bygninger, veje til hele byer kan vi således repræsentere i et mere praktisk forhold - eksempelvis størrelsesforholdet 1:800. Træer bliver til cirkler, bygninger til polygoner og gader til tykke linjer. Ønsker vi bedre overblik sigter vi højere - i 1:500.000 bliver større byer til små cirkler, motorveje til tynde linjer og andre ting forsvinder helt. Dette er netop hvad begrebet semantisk zoomning dækker over - repræsentationsændring i forhold til det niveau hvorfra en given scene anskues.

Begrebet semantisk zoomning blev for første gang introduceret i 1993 af K. Perlin og D. Fox i Pad-systemet [1] og herefter senere yderligere uddybet og benyttet i Pad++ af B. Bederson og J.D. Holland [2]. Idéen udsprang fra et ønske om at give brugerene optimale forhold til at kunne udnytte den tilrådighedstillede arbejdsplads. Det foresloges derfor at introducere en grænseflade hvor interaktion på ekstremt fint niveau er muligt - altså et virtuelt uendeligt arbejdsrum. Som Perlin og Fox skriver

*"Imagine that the computer screen is a section of wall about the size of a typical bulletin board or whiteboard. Any area of this surface can then be accessed comfortably without leaving one's chair. Imagine further that by applying extraordinary good eyesight and eye-hand-coordination, a user can both read and write as comfortably on any micron wide section of this surface as on any larger section. This would allow for full use of a surface which is several million pixels long and high, on which one can comfortably create, move, read and compare information at many different scales" [1] pp. 1.*

Selvfølgelig er ovenstående scenario praktisk umuligt. Vi kan ikke naturligt interagere med objekter på sådanne mikroskopiske skala. Den naturlige konsekvens heraf var da at introducere et system som efterligner denne funktionalitet - nemlig et system i form af et 'magisk forstørrelsesglas'. Herved ville det være muligt, ved hjælp af zoomning, at konceptualisere et sådanne uendeligt arbejdsrum.

I sig selv var denne idé ganske interessant, men den opfyldte ikke hele den vision som Perlin og Fox havde. Derfor introduceres begrebet semantisk zoomning. I modsætning til almindelig geometrisk zoomning - hvor et givent objekt blot skales på almindelig vis - som når man kører i bil og ser skilte, bygninger og andre ting i det fjerne, blive større og større - tilføjer semantisk zoomning en ændring i objektets repræsentation i forhold til en

given skalering. Der er altså tale om eksempelvis ændringer i størrelse, form, farver og detaljeringsniveau, hvorved målet - at brugeren komfortabelt skal kunne skabe, flytte, læse og sammenligne informationer i mange forskellige skaleringer - bliver opfyldt.

## 1.1 Anvendelsesmuligheder

Reelt set er der et utal af situationer hvor princippet om semantisk zoomning kan benyttes, og princippet er da også almindeligt brugt i enkelte tilfælde - eksempelvis, som nævnt ovenfor, i kartografien. Om end det hovedsageligt kun er i denne forbindelse, at metoden har fundet fast indpas findes princippet grundlag, den geometriske zoomning i utallige af grænseflader - alle under betegnelsen zoombare grænseflader. Der er dog mange forskningsprojekter, inden for menneske-datamaskine interaktion, som har taget emnet semantisk zoomning op. Disse projekter spænder over en række meget forskellige områder. Eksempelvis kan nævnes projekter med semantisk zoomning benyttet på større billedsamlinger [3], tekst [4], videoredigering [5] og dokumentsamlinger [6].

Af projekter som er relevante i forhold til nærværende opgave findes Table Lens [12]. Peter Piroll og Ramana Rao har her udviklet og testet et visualiseringsværktøj til at søge efter mønstre og afvigelser i multivariate dataark. Om end deres metode - en blanding af tabeller med grafiske repræsentationer og et antal muligheder for direkte manipulation - ligger et stykke fra metoderne i nærværende projekt så er det væsentligt at nævne eftersom det peger i retning af at konceptet også er overvejet på almindeligt numerisk materiale.

Der findes dog umiddelbart ikke projekter som behandler metoderne foreslået i dette projekt - altså den semantiske zoomning benyttet på en tabular fremstilling af større statistiske dataark.

## 1.2 Generelle problemstillinger indenfor semantisk zoomning

Om end den gængse definition af zoomning vil være et spørgsmål om en ændring i afstanden et given objekt ses fra, kan begrebet også bruges på anden vis. Herved kan zoomning også foretages over andre egenskaber som eksempelvis tid (hvordan ændrer befolkningstætheden, angivet på et kort, sig over tid) og pris (hvilke lejligheder er tilgængelige til et givent prisniveau) - hvorved ændringen ikke ses i afstanden fra objekterne men i tætheden eller arten af disse. Som angivet i afsnit 2.1 og 3 er der i denne opgave dog fokuseret på zoomning jf. den gængse definition. Det vil således være afstanden til tabellen som angiver zoomniveauet.

## 2 Problemformulering

*“I hvilken grad støtter princippet om semantisk zoomning brugere i deres arbejde med indledende trendanalyse og delperiodeopdeling af større statistiske dataark repræsenteret i regnearksform.”*

### 2.1 Afgrænsning

Denne opgave vil være afgrænset til at evaluere i hvilken grad princippet om semantisk zoomning støtter brugere i deres arbejde med indledende trendanalyse og delperiodeopdeling af større statistiske dataark repræsenteret i regnearksform. Hvad der menes med 'støtte brugeren' er nærmere uddybet i afsnit 7.1 på side 28.

Det antages her at der er tale om statistiske dataark som udvikler sig over tid. Det er altså rækkernes udvikling over kolonner som vil være af interesse. Arbejde er her nærmere afgrænset til værende det indledende arbejde som en given bruger nødvendigvis vil kunne tænkes at foretage sig på materialet. Dette værende sig eksempelvis et ønske om at danne sig et overblik over dataene, altså udtrække generelle tendenser, have mulighed for at gå i dybden med enkelte delområder og se indbyrdes forhold over flere rækker. Der er altså derfor ikke tale om en evaluering af en brugers dybdegående arbejde og manipulation med disse data, men blot den indledende fase - også kendt under begrebet udforskende dataanalyse. Disse opgaver vil blive nærmere defineret i afsnit 3 på side 12.

Opgaven vil endvidere kun fokusere på disse dataark repræsenteret i tabelform. Der vil således ikke være tale om udvikling af metoder til brug på diagrammer og lignende repræsentationer af de givne data - dette skyldes der på dette område allerede eksisterer mange studier, hvorfor det er ønskeligt at udforske en ny og alternativ måde at visualisere sådanne data.

For at kunne vurdere hvorledes semantisk zoomning støtter brugeren mere end konventionelle løsninger er brugertestene blevet udført på såvel det udvalgte løsningsforslag, beskrevet i afsnit 5 på side 21, som en referenceløsning som kun benytter sig af geometrisk zoomning.

Overordnet set vil denne opgave altså indeholde følgende delopgaver

- Redegørelse for begrebet semantisk zoomning
- Identificering af problemer med indledende analyseopgaver ifm. større statistiske dataark.
- Udarbejdelse af et forslag til en løsning hvori semantisk zoomning bruges til at afhjælpe ovenstående problemer
- Implementering af udvalgt løsningsforslag i et prototypeprogram

- Evaluering af det implementerede løsningsforslag samt referenceløsningen, herunder
  - Redegørelse for evalueringsmetode
  - Undersøgelse af implementerede løsningsforslag og referenceløsning
  - Evaluering af undersøgelse

Der vil ikke blive lagt vægt på implementationernes robusthed, dybdegående funktionalitetsbeskrivelser eller programmeringsmæssige overvejelser og beskrivelser af løsningsforslagene.

## 2.2 Motivation

Motivationen for denne opgave er mit personlige studieliv. Jeg har således i forbindelse med et sidefag i økonomi, ved Økonomisk Institut, Københavns Universitet, stiftet bekendtskab med de økonomistuderendes opgaver i dagligdagen. Især i kurset Samfundsbeskrivelse - en indførelse i deskriptiv statistik og det danske samfunds væsentlige institutioner og mekanismer - har det været muligt at føle hvor vigtig en rolle edb spiller også i denne faggruppe. Kurset Samfundsbeskrivelse baseres i høj grad på en række literære værker som er kraftigt statistisk funderet - herved menes at næsten alle pointer belyses ved relevant samfundsstatistik - eksempelvis nationalregnskabet, fertilitetstal og lønindeks. Det er også disse kilder som er fokus i de obligatoriske opgaver, samt eksamen, hvis mål er at give de studerende praktisk erfaring med at arbejde med disse statistiske data - samtidig med at den tilhørende teori hæftes sammen hermed. En stor del af de studerendes arbejde er derfor placeret i arbejdet med disse data - et arbejde som typisk udføres i regnearksmiljøer som Microsoft Excel<sup>1</sup> og PC-AXIS<sup>2</sup> - som de studerende introduceres for i undervisningen. Disse applikationer er dog begrænsede i deres forhold til at præsentere statistisk data. Således er det den almindeligt kendte regnearksmodel som benyttes - dog med mulighed for simpel geometrisk zoomning. Motivationen for denne opgave ligger i netop denne begrænsning. Tanken er således, at det må være muligt at finde alternative metoder til at præsentere sådanne data - metoder hvoriblandt semantisk zoomning eksempelvis findes. Det er da målet at undersøge om denne teknik kan byde på et relevant alternativ til den gængse metode.

---

<sup>1</sup>Microsoft Excel er en regnearksapplikation som udvikles og vedligeholdes af Microsoft Corporation.

<sup>2</sup>PC-Axis er en applikation til at præsentere statistisk data. PC-Axis udvikles og vedligeholdes af Statistiska Centralbyrån som er det svenske modstykke til Danmarks Statistik.

### 3 Identificering af opgaver

Som beskrevet i afsnit 2.2 er motivationen for denne opgave kraftigt funderet på de erfaringer der gøres igennem kurset Samfundsbeskrivelse ved Økonomisk Institut på Københavns Universitet. Det må således betragtes som legalt at basere væsentlige antagelser om datanes art på det materiale som de studerende sættes overfor i dette kursus.

Helt specifikt skal de studerende gennem kurset løse fire obligatoriske hjemmeopgaver og en eksamensopgave. Hjemmeopgaverne og Tag-Hjem-delen af eksamensopgaven fastsættes oftest på formen "*Beskriv og forklar udviklingen i X siden Y*" - eksempelvis kan nævnes følgende

*"Beskriv og forklar udviklingen i udvalgte prisindeks siden 1983"*  
- Eksamen i faget samfundsbeskrivelse, 1. år, økonomisk kandidatexamen, Københavns Universitet, Økonomisk Institut, januar 2006

*"Beskriv og forklar udviklingen i betalings- og kapitalbalancen siden 1991 og så langt frem som muligt"* - Eksamen i faget samfundsbeskrivelse, 1. år, økonomisk kandidatexamen, Københavns Universitet, Økonomisk Institut, januar 2001

Af primære statistiske kilder benyttes da materiale fra Danmarks Statistiks online publikationer såvel som trykte værker. Et dataark som ofte benyttes er danmarks nationalregnskab[10]. Dette dataark, bestående af 3.584 værdier<sup>3</sup>, stiller den studerende overfor et regneark som er uoverskueligt og svært at tolke. Dette er eksempelvis kun ét af en lang række dataark som de studerende skal berøre på en eller anden vis i forbindelse med løningen af en sådanne opgave.

Det vil således være denne type dataark og opgaver, som de studerende ved økonomistudiet på Københavns Universitet stilles overfor i faget samfundsbeskrivelse, som vil danne grundlaget for de antagelser omkring opgaver og typer af data som behandlingen i denne opgave vil baseres på.

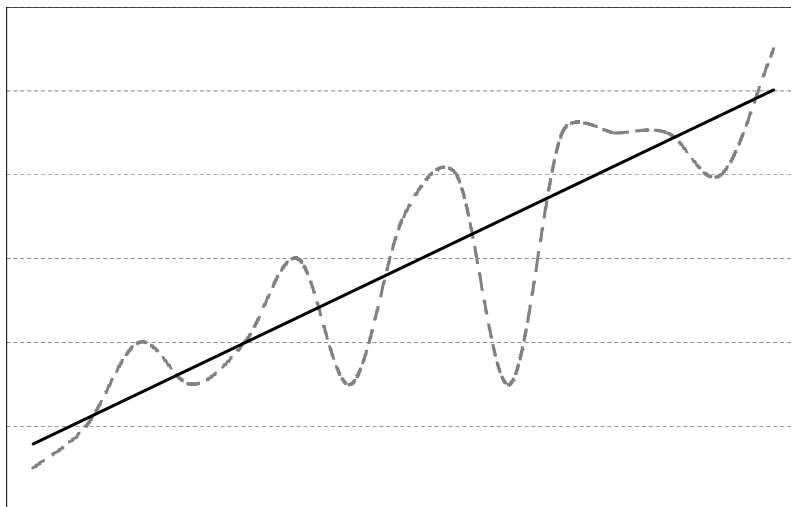
#### 3.1 Opgavedefinitionen

Indledningen ovenfor angiver de typiske opgaver som stilles i faget Samfundsbeskrivelse. Hertil er to delelementer - en beskrivende del og en forklarende del - hvoraf den beskrivende del er den del som søges støttet ved metoderne i nærværende opgave.

Vurdering af disse delopgaver sker jf. de udarbejdede rettevejledninger. Heraf fremgår det f.eks. at der lægges vægt på at

---

<sup>3</sup>56 konti og 64 tidspunkter, én prisenhed og sæsonkorrigering valgt over perioden 1990 til 2005 (kvartalsvise værdier)



Figur 1: Dataserie med angivelse af overordnet trend

*“Udviklingen beskrives kort med fokus på overordnede trend samt en opdeling i delperioder.”* - Rettevejledning til tag-hjem opgaven, 2. del af eksamen i faget samfundsbeskrivelse, 1. år, økonomisk kandidateksamen, Københavns Universitet, Økonomisk Institut, januar 2005

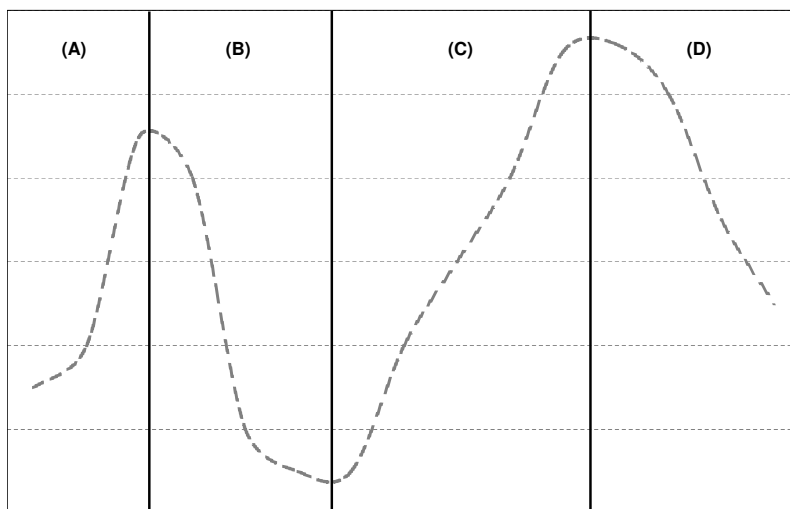
De væsentlige pointer har altså fokus på overordnet trend og opdeling i delperioder.

### 3.1.1 Overordnet trend

Ved overordnet trend menes fokus på bevægelser i dataene på langt sigt. Således vil mindre fluktuationer fra den generelle tendens være uinteressant. Der er altså tale om et ønsket fokus på udviklingen over hele perioden - dette værende sig for opadgående, for nedadgående eller stagnerende tendenser. Dette kan beskrives ved eksempelvis total absolut ændring eller gennemsnitlig årlig vækstrate - det vigtige er at det klart og entydigt fremgår i hvilken retning de givne data bevæger sig. Figur 1 illustrer tanken bag den overordnede trend. Således er den sorte linje altså en simpel afbildning af seriens overordnede trend. Som det ses belyser den overordnede trend altså niveauskiftet i dataene fra starten af perioden til slutningen af perioden samt evt. den hast hvorved udviklingen finder sted.

### 3.1.2 Opdeling i delperioder

Hvorimod spørgsmålet om den overordnede trend umiddelbart må siges alene at være et spørgsmål om beregning og valg af repræsentation ligger der



Figur 2: Dataserie opdelt i delperioder

mere i opgaven omkring opdeling i delperioder. Der er altså her nærmere tale om et analyseproblem frem for et simpelt beregningsproblem. Ved et givent dataark skal opgaveløseren altså finde intervaller med lignende udviklingstendenser. Man kan umiddelbart sige at princippet fra den overordnede trend skal benyttes på delintervaller af perioden således at disse delintervaller udviser samme trend. Det vil således være naturligt at opdele en given periode i delperioder hvor den givne dataserie ændrer udvikling fra værende for opadgående til for nedadgående eller fra kraftig vækst til svag vækst.

Som det ses af figur 2 er dataserien her opdelt i fire delperioder -A, B, C og D. Delperiode A og C med for opadgående trend og delperiode B og D med for nedadgående trend.

### 3.2 Antagelser om datatypen

Som også skrevet i afgrænsningen på side på side 10 er en væsentlig antagelse om dataenes natur at der er tale om rækker der udvikler sig over tid. Således er der altså tale om numeriske tabeller med en titel til hver række samt en tidsangivelse for hver søjle, og ikke krydstabeller med andre slags data som eksempelvis *'fertiliteetskoefficient fordelt på mødrenes uddannelsesniveau og bopæl'*. Alle rækker vil derfor være på formen *'Serie X's (række) udvikling over perioden A - B (kolonner)'*.

Om end dette er en kraftig begrænsning af typen af data må dette dog anses som ganske naturligt og legalt i forhold til opgaveidentifikationen i indledningen til nærværende afsnit.

Af opgaveidentifikationen kan da identificeres et par klart overordnede klasser af data. Dette værende sig

a1	a2	a3
b1	b2	b3
c1	c2	c3
d1	d2	d3
e1	e2	e3

Figur 3: Referenceløsning

1. Tæt knyttede serier
2. Ikke direkte sammenhængende serier

**Ad 1** Herved menes serier (rækker) som er direkte relaterede - eksempelvis delkomponenter af en samlet serie. Nationalregnskabet er et udemærket eksempel på dette - her er flere overordnede konti yderligere uddybet i underkonti. Således er der tale om en hierarkisk struktur med naturlige grupperinger.

**Ad 2** Ved 'ikke direkte sammenhængende serier' menes serier som der på forhånd ikke kan angives nogen direkte relation imellem. Dette værende sig forskellige serier som 'bruttoværditilvækst' og 'salg af varige forbrugsgoder'. Ved angivelsen 'ikke direkte' understreges det at der til trods for en mulig eksisterende sammenhæng - som i nævnte eksempel - er denne sammenhæng ikke kendt. Der er altså blot tale om en række serier uden indbyrdes struktur.

## 4 Metoder til semantisk zoomning på statistiske dataark

Af figur 3 ses den tabulare referenceløsningen, som det f.eks. kendes fra Microsoft Excel og lignende. Der er altså i princippet blot tale om  $n \times m$  matrice. Hertil kommer dog også en række med angivelse af tidsenheder samt en kolonne med angivelse af rækkenavne - hvilket i realiteten vil være en  $(n + 1) \times (m + 1)$  matrice, men hvoraf det interessante umiddelbart kun er delen  $n \times m$ .

I det følgende vil blive beskrevet et udvalg af metoder til semantisk zoomning over denne  $n \times m$  referenceløsning.

### 4.1 Generelt

Målet med semantisk zoomning i dette tilfælde er at give brugeren et bedre overblik over dataene end det opnås med den gængse geometriske zoomning. Hertil er der især et enkelt generelt princip som kan gøres brug af - reducere af den mængde objekter brugeren præsenteres for. [18] angiver at rod

i den visuelle repræsentation kan have negative virkninger lige fra aftagende brugseffektivitet til at grænsefladen er mindre appellerende for brugeren. Derfor må det være et mål i sig selv at forsøge at begrænse den mængde data som brugeren præsenteres for.

En anden vigtig generel faktor må være genkendelse. Brugere skal støttes i at have mulighed for at genkende det, der præsenteres. Dette vil i praksis betyde at brugeren altid skal stilles overfor samme repræsentation under samme forhold men på forskellige tidspunkter - der må altså ikke være arbitrære dynamisk beregnede repræsentationer på forskellige områder af arket og forskellige zoomniveauer. Brugeren skal med andre ord have mulighed for at vende tilbage til data denne allerede er blevet præsenteret for én gang.

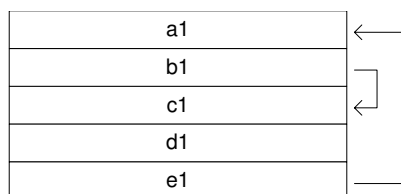
## 4.2 Gruppering

En triviel metode til semantisk zoomning på dataene er idéen om gruppering. Herved menes en enten vertikal eller horisontal aggregering af et antal rækker eller søjler. Der kan i dataene således eksistere naturlige grupperinger - set i eksempelvis nationalregnskabet, hvor enkelte rækker er underposter til en overordnet række. Disse kunne da samles og repræsenteres ved blot én række. Over søjler kunne man forestille sig en naturlig gruppering i form af de benyttede tidsenheder. I fald der i dataene er en søjle pr. kvartal ville en naturlig gruppering da være over år. Således at kvartaler i et givent år aggregeres til blot ét år og derved kun én søjle.

Det er dog vigtigt at bemærke at denne metode ikke er generisk. Herved menes at dataene skal indeholde informationer om sådanne mulige grupperinger for at det kan give mening. Naiv gruppering ville kunne betyde at essentielle data aggregeres med andre data, og derved reelt forsvinder for brugeren. Endvidere kunne det risikeres at der grupperes meget uhensigtsmæssigt - som i eksemplet med nationalregnskabet - kunne private konti blive uheldigt grupperet med offentlige konti.

## 4.3 Filtrering

En teknik, som umiddelbart kan minde om idéen om gruppering, er idéen om filtrering. For at undgå at overflødige og statistiske insignifikante data oversvømmer repræsentationen kan data filtreres fra. Denne metode vil umiddelbart, i henhold til antagelserne omkring typen af data som denne opgave omhandler, kun være relevant udført over rækker. Man kunne således forestille sig at visse rækker fjernes fra en given repræsentation. Herved ville data som er uinteressante i helhedsbilledet - nulrækker eller uvæsentlige delkomponenter - skjules hvorved repræsentationen indeholder færre elementer. Man skal dog være opmærksom på at denne teknik, ligesom ved gruppering, vil være uhensigtsmæssig at bruge i forbindelse med ikke direkte sammenhængende serier da en eventuel ønsket sammenligning over rækker



Figur 4: Sortering af rækker

kan være forhindret ved at de enkelte sammenligningselementer er filtreret væk.

#### 4.4 Sortering

Om end der ikke umiddelbart er gjort nogle antagelser om dataenes organisering i forhold til hinanden - andet end skelnen mellem tæt knyttede og ikke direkte sammenhængende serier - så er netop denne faktor også manipulerbar. Ved at ændre seriernes indbyrdes organisering, eksempelvis ved sortering efter vægte som rækkens relative vægt i forhold til hele dataarket eller størrelsen på udviklingen over rækken, kunne det tænkes at brugeren vil have nemmere ved at finde relevante rækker. Det skal bemærkes at, jf. antagelserne om dataene, at sortering over søjler ikke vil give mening.

Som det ses af figur 4 så er der altså blot tale om en ændring i forhold til rækkens placering og ikke rækkens repræsentation. Det kunne tænkes at relevante rækker herved placeres øverst hvorved brugeren sparer tid på at finde frem til netop disse rækker.

#### 4.5 Generelt om sammenfletning af celler

Som kort nævnt tidligere kan en generel teknik, til semantisk zoomning på sådanne tabeller, tænkes at være idéen om sammenfletning og aggregering af celler. Herved forstås altså at et antal celler, over rækker eller over søjler, flettes sammen til blot én.

##### 4.5.1 Vertikal sammenfletning

Den vertikale sammenfletning er kort omtalt i afsnittet om gruppering. Der er altså tale om én eller flere rækker som slås sammen med en anden række hvorved antallet af objekter mindskes.

I figur 5 på den følgende side ses den vertikale sammenfletning som blot en opskalering af én rækkes højde. Arealet af de celler der flettes sammen ændres derfor ikke. Om end dette kun vil give mening i fald rækkerne ligger umiddelbart under hinanden er det værd at bemærke at denne metode vil gøre det lettere for brugeren at genkende det enkelte område. Som nævnt i afsnit 4.1 er det vigtigt at bevare en given repræsentation på et givent

a1	a2	a3
d1	d2	d3
e1	e2	e3

Figur 5: Eksempel på vertikal sammenfletning af celler

zoomniveau. Så for at vertikal sammenfletning, med rækker ikke liggende i nærheden af hinanden, vil der være behov for en sortering af rækkerne forinden dataarket vises.

Idéen om vertikal sammenfletning må dog umiddelbart skulle bruges med varsomhed. Som nævnt tidligere vil det kun være i forbindelse med naturlige grupperinger at teknikken reelt kan bruges hensigtsmæssigt.

I fald teknikken bruges rejser da et spørgsmål om sammenfletning på dele af rækker giver mening. Umiddelbart kan man forestille sig scenarier hvor dette ville kunne fungere ganske udemærket. I fald man eksempelvis har to brudte dataserier som begge afslutter på samme tidspunkt, hvorefter de fortsætter som nulrækker, vil de to rækker kunne sammenflettes på blot den sidste del uden problemer. Umiddelbart skal man dog være påpasselig med at forsøge dette i rækker med egentlige informationer da dette ville kunne tænkes at bryde brugerens skimmen hen over en given række. Endvidere ville det kunne tænkes at store sammenflettede områder, midt i en række mindre områder, ville tiltrække unødigt opmærksomhed. Brugeren kunne have svært ved at abstrahere fra at der muligvis blot var tale om et område med lignende uinteressant udvikling i forhold til et vigtigt område repræsenteret ved størrelsen.

#### 4.5.2 Horisontal sammenfletning

I figur 6 ses et eksempel på horisontal sammenfletning af celler. Under antagelserne om dataenes art er der intet problem i denne form for sammenfletning. Denne metode giver en lang række muligheder for manipulation med og beregninger på dataene. Da sammenflettede celler udgør et sæt af værdier kan en række forskellige tal beregnes - eksempelvis årlig gennemsnitlig vækstrate, median, mm. - mere herom i afsnit 7 til 12.

Et spørgsmål som dog rejser sig er hvorvidt scenariet i figur 6 bør forekomme eller ej. Man kunne forestille sig at sammenfletningen kun bør forekomme i fald at søjlesider ikke brydes. I scenariet i figuren kan det tænkes at brugerens læsning af en søjle kan blive afbrudt af en sammenflettet række celler som har den givne søjle i dens spændvidde. Dette ville betyde besvær med sammenligning af rækker på bestemte tidspunkter i serien.

a1		a3
b1		b3
c1	c2	
d1		
e1	e2	e3

Figur 6: Eksempel på horisontal sammenfletning af celler

#### 4.6 Repræsentationer på celleniveau

Afhængigt af hvilken overordnet metode der vælges, vil der være en række repræsentationsmetoder på celleniveau som giver mening - disse præsenteres i det følgende. De enkelte repræsentationsmetoder vil kun give mening i forbindelse med specifikke overordnede metoder. Således vil en angivelse af den relative ændring i celler der er flettet over flere rækker ikke give mening - eftersom der i denne opgave er tale om udvikling over tid over kolonner, jf. afsnit 3.2 på side 14. Ligeledes vil aggregering over rækker ikke give mening i fald der er tale om rækker uden direkte sammenhæng.

Det er selvfølgelig være muligt at kombinere disse repræsentationsmetoder indbyrdes, således ville en angivelse af et interval kunne følges med en angivelse af den relative ændring, eller middelværdien.

**Referenceløsning** Af figur 7 ses referenceløsningen på celleniveau. Her er altså tale om en celle indeholdende en numerisk værdi. Med hvilken enhed, om der er kr, grader, antal eller andet, eller hvilken skala, tusinde, millioner, tiendedele eller andre, vides ikke på forhånd.

a1
----

Figur 7: Cellerepræsentation, referenceløsning

**Absolutte ændringer** I fald der er tale om en celle, som er et resultat af en aggregering, kan cellen repræsenteres som den absolutte ændring, der er over de celler den aggregerer. Således kan cellerne med værdierne 124.100, 130.050, 135.875, 147.350 repræsenteres, som i figur 8, ved blot 23.250 - altså den absolutte ændring fra starten af serien til slutningen af serien.

23.250
--------

Figur 8: Cellerepræsentation, absolut ændring

**Relative ændringer** De absolutte ændringer, beskrevet i afsnittet ovenover, kan imidlertid også repræsenteres på anden vis - nemlig ved relative

ændringer. Her er altså tale om en repræsentation som ikke er biased imod værdiernes absolutte niveau, men som blot udtrykker den relative ændring over serien.

2,1 %
-------

Figur 9: Cellerepræsentation, relativ ændring

Sådanne relative ændringer kan, som i figur 9, repræsenteres ved et udtryk for den procentvise ændring. Udført over en hel række kan ændringerne enddog repræsenteres ved indekstal hvorved udviklingen i sin helhed udtrykkes klart. Også udtryk for den gennemsnitlige vækstrate over en serie kunne tænkes.

**Enhedsskift** En meget generel repræsentationsændring er idéen om enhedsskift. Ved at angive værdierne i tusinde, millioner eller tiendedele kan unødvendige cifre spares. 23.000 kan således, som i figur 10, blive til 23 ved enhedsskift til tusinde.

23
----

Figur 10: Cellerepræsentation, enhedsskift

Denne metode kan, i sin generelle natur, bruges på såvel aggregerede celler som singulære celler.

**Intervaller** Idéen om intervaller er ganske naturlig. Ved at repræsentere start og slut på en given aggregeret serie gives et indtryk af såvel udviklingen som de absolutte værdier denne serie indeholder.

43 - 52
---------

Figur 11: Cellerepræsentation, interval

Af figur 11 ses én måde at angive intervallet på - umiddelbart den mest simple. Andre repræsentationer kan f.eks. være  $[a - b]$  og  $[a; b]$ .

**Datainformationer** En metode der helt fjerner sig fra de oprindelige værdier er idéen om repræsentation ved datainformationer. Ved datainformationer forstås beregnede repræsentationer af de oprindelige værdier. Dette eksempelvis værende sig beregnet middelværdi, standardafvigelse, ekstremaer,

vægtede gennemsnit<sup>4</sup>, vækstbidrag<sup>5</sup>, differentialkvotienter og andre beregnede størrelser.

M = xx, SD = yy
-----------------

Figur 12: Cellerepræsentation, informationer

Denne metode kan i princippet bruges over både kolonner og rækker, men som nævnt vil nogle af datainformationerne umiddelbart give mest mening i én retning. Som det ses af figur 12 kan disse datainformationer sammenstilles i en given celle. Det vil dog være vigtigt at der vælges klare tekster til angivelse af de enkelte informationer. Endvidere skal man være opmærksom på at et forsøg på at sammenstille for mange informationer i en aggregeret celle højst sandsynligt vil betyde at fordelene ved den semantiske zoomning aftager.

## 5 Beskrivelse af løsningsforslag

I det følgende vil blive gennemgået og argumenteret for det udvalg af metoder som udgør det samlede løsningsforslag som ønskes evalueret. Beskrivelsen af selve applikationen vil blive gennemgået i afsnit 6.

### 5.1 Overordnede metoder

For at søge en minimering af den mængde data brugeren præsenteres for er det valgt at implementere idéerne om vertikal og horisontal sammenfletning af celler. Løsningen sørger da for fravælgelse af rækker, jf. afsnit 4.3, samt sammenfletning af kolonner, jf. afsnit 4.5 og 6.

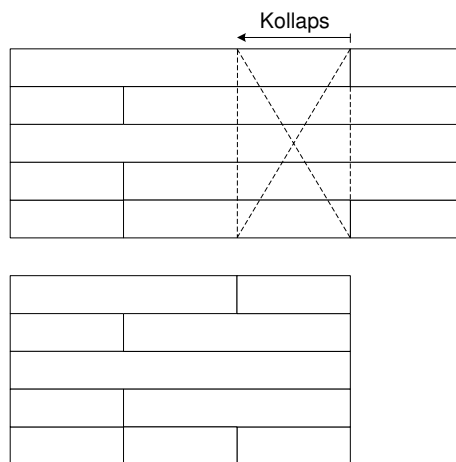
#### 5.1.1 Rækkeoperationer

Det er valgt at implementere en naiv frasortering af rækker. Således bliver rækker som vægtes insignifikante, i forhold til de andre rækker i arket, sorteret fra. I afsnit 5.2.2 beskrives det tekniske i metoden. Dette princip er valgt for betydeligt at reducere antallet af celler på lave zoomniveauer. Frasortering af rækker sker automatisk, men med mulighed for at brugeren udvælger rækker som ikke ønskes frasorteret. Dette er for at sikre at brugeren

---

<sup>4</sup>Dette giver umiddelbart mest mening over rækker - således kan et felt fra en række indregnes i gennemsnittet over flere rækker med rækkens relative vægt i forhold til alle rækkerne. Sagt på en anden måde - hvis rækkerne  $1 \dots n$  har vægtene  $v_1 \dots v_n$  vil en celle,  $x_i$  indgå i det vægtede gennemsnit med  $v_i \cdot x_i$ .

<sup>5</sup>Igen giver dette umiddelbart kun mening over rækker - særligt kun i tilfældet hvor der er tale om tæt knyttede serier. Således kan rækker sammenflettes og aggregeringens vækstbidrag til det overordnede komponent kan beregnes.



Figur 13: sammenfletning af kolonner. Øverst ses dataarket forinden søjlen sammenflettes og nederst efter kollapset.

får udbytte af de andre principper i løsningen på lave zoomniveauer til trods for at rækken ville være sorteret fra forinden.

### 5.1.2 Kolonneoperationer

Som nævnt er det valgt at implementere princippet omkring sammenfletning og gruppering af kolonner. Jf. afsnit 3 er det opgaver i analyse af overordnede trends samt opdeling i delperioder som ønskes støttet. Derfor er der implementeret en løsning som, rækkevis, sammenfletter kolonner - som det ses i figur 6 på side 19. Således mindskes antallet af data jo længere der zoomes ud. For at forhindre at hele rækker sammenflettes til én celle med ét er det valgt at begrænse længden på en sammenfletning i forhold til zoomniveauet. Således bliver sammenfletninger længere ved lavere zoomniveauer og en forholdsvis flydende overgang fra enkeltcellerepræsentation til den endelige repræsentation opnås. For at finde frem til hvilke celler som er interessante i et udviklingsforløb får hver celle tilknyttet en interessegrad. Denne interessegrad udtrykker en given celles placering i forhold til en igangværende udvikling samt om cellen medfører et brud i udviklingen - eksempelvis fra for opadgående til for nedadgående tendens. Den tekniske del af denne implementation er gennemgået i afsnit 5.2.1. Denne interessegrad benyttes til at bestemme hvornår en given serie skal brydes op.

Udover denne sammenfletning af celler er også implementeret egentlig sammenfletning af kolonner. Således bliver kolonner som alle blot er et delinterval af andre intervaller i alle rækkerne, sammenflettet - dette er illustreret i figur 13.

Dette princip er valgt implementeret for at forhindre lange ensartede sammenfletninger af celler fra at tage mere skærmpads end nødvendigt.

Søjleoverskrifterne ændres da blot til at være et interval der angiver start- og sluttidspunkt for en gruppe af sammenflettede søjler. Princippet er også vist i figur 16.

### 5.1.3 Celleoperationer

I implementationen er det valgt at benytte skiftende repræsentationer alt efter zoomniveau. Der er udvalgt følgende repræsentationsmetoder, med øverste angivende repræsentationen ved 100% skalering og derefter ved nedskalering.

- Oprindelige værdier med tusindtalsseparatorer
- Intervaller angivet i større enheder - således bliver millioner til milliarder
- Angivelse af gennemsnitlig årlig vækstrate og medianen for sammenfletningen
- Angivelse af gennemsnitlig årlig vækstrate

Tanken er således at informationsmængden mindskes jo længere der zoomes ud. Intervallerne kan undlade flere mellemliggende værdier og enhedsskiftet sørger for at de væsentlige informationer i værdierne udtrækkes. Angivelsen af den gennemsnitlige årlige vækstrate er et kort og koncist udtryk for udviklingen over intervallet, og medianen hjælper med til at sætte fokus på de reelle værdier bag. Det yderste niveau udtrykkes blot med angivelse af den gennemsnitlige årlige vækstrate for at sikre et nemt overblik over de generelle tendenser i dataarket.

## 5.2 Overgange mellem repræsentationsformer

Et spørgsmål som naturligvis rejste sig under udviklingen af løsningsforslaget var hvordan overgange mellem repræsentationsformer skulle foregå. I det følgende er en beskrivelse af de mekanismer som sørger for repræsentationsskift.

Et generelt spørgsmål er dog den reelle overgang mellem to givne repræsentationsformer. Her findes umiddelbart to overordnede muligheder - glidende overgange og pludselige overgange. I udviklingsforløbet blev begge metoder udviklet og vurderet. Til trods for det umiddelbart tiltalende i glidende overgange blev det fravalgt i den endelige løsning. Korte pilotafprøvninger af implementationen med glidende overgange<sup>6</sup> viste en klar problematik i denne form for overgang. Den megen aktivitet på skærmen fra animering af celler til højre og venstre og rækker og søjler som sammenflettes virkede mere forvirrende og forstyrrende end nyttig. Endvidere viste det sig at brugeren

---

<sup>6</sup>Der blev implementeret skeduleringsmekanismer til styring af al nødvendig animation af celler for tilpasning til en given repræsentation i en given række samt animering af sammenfletning af søjler og rækker.

sjældent ville zoome langsomt ind og ud hvorfor animationen kun havde et meget begrænset tidsrum at blive udført i hvis ikke brugeren skulle blive generet af stop i skaleringen. Derfor er den endelige implementation baseret på pludselige repræsentationsskift.

### 5.2.1 Interessegrad for celler

Den udvalgte metode til sammenfletning af celler i rækker er baseret på en række egenskaber nærmere uddybet i det følgende. Hver celle får tilknyttet en interessegrad<sup>7</sup> som danner grundlaget for hvornår celler flettes og ej. Interessegraden defineres som følger.

$$DOI_n = \left| \frac{V_n - G}{G} \right| \cdot Z$$

hvor  $DOI_n$  er interessegraden for celle  $n$ ,  $V_n$  er værdien i celle  $n$  og

$$G = \frac{\sum_{i=n-1}^{n+1} V_i}{3}$$

og  $Z = 1$  hvis

$$\frac{(V_n - V_{n-1})}{|(V_n - V_{n-1})|} = \frac{(V_{n+1} - V_n)}{|(V_{n+1} - V_n)|}$$

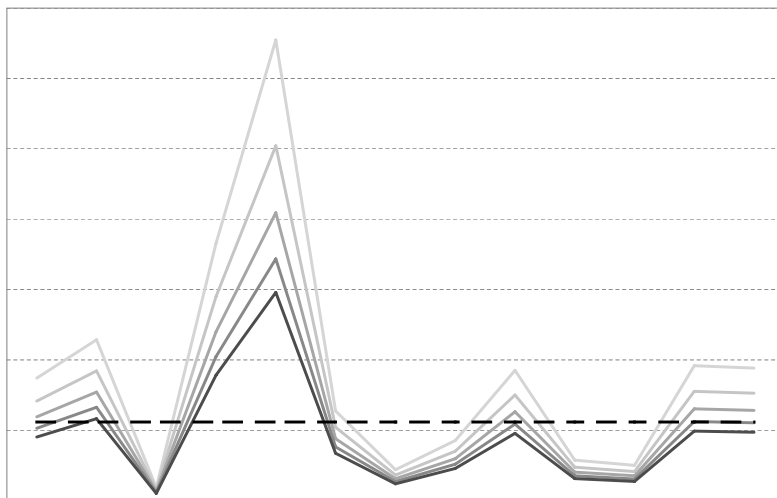
og ellers er  $Z = 1, 1$ . Med andre ord - interessegraden er bestemt af cellens afvigelse fra gennemsnittet af de to omkringliggende celler samt sig selv, dette i forhold til samme gennemsnit, og endeligt skaleret med yderligere 10 % hvis fortegnet af ændringen fra cellen mod venstre ikke er lig med fortegnet af ændringen fra cellen mod højre. I tilfældet hvor en af ændringerne er 0 betragtes det også som at ændringerne ikke er ens. Herved opnås et udtryk for cellens betydning i en igangværende tendens.

**Sammenhæng mellem cellers interessegrad og zoomniveau** Denne interessegrad bruges da til at bestemme en rækkes opdeling på et givent zoomniveau. Dette sker på følgende måde.

$$DOI_n^s = 2^{s \cdot (1/DOI_n)}$$

Med andre ord interessegraden for en given celle  $n$  på et givent zoomniveau  $s$ , altså  $DOI_n^s$ , er to opløftet i cellens inverse interessegrad skaleret med det givne zoomniveau. Interessegraden vil altså blive belønnet mere jo længere der er zoomet ud jo højere den initielt var. Grænseværdien for hvornår en celle er interessant ved et givent niveau er fastsat på følgende vis

<sup>7</sup>På engelsk: degree of interest (DOI)



Figur 14: Interessegraden udtrykt ved forskellige zoomniveauer samt grænseværdi angivet ved stiplede linje

$$Grnse_{DOI} = 4^{2T}$$

hvor  $T$  er angivet som standardafvigelsen på gennemsnittene af alle de zoomniveauspecifikke interessegrader ( $\sum_{i=1}^n DOI_i^s$ ). Dette kan give et billede som udtrykt i figur 14.

Som det ses vil på forhånd interessante celler sørge for at resultere i et brud i sammenfletningen til trods for at der zoomes ud mens mindre interessante celler resulterer i brud alt efter hvilket zoomniveau der er tale om.

For at undgå at hele rækker sammenflettes til én celle allerede i første repræsentationsændring er der defineret en grænse for en sammenfletnings maksimale spændvidde på et givent zoomniveau. Dette er fastsat ved,

$$L_s = \left\{ \left( \frac{1}{s} \right)^8 \right\} - 1$$

Altså, den maksimale længde af en sammenfletning af celler på et zoomniveau  $s$ , altså  $L_s$ , er den inverse til  $s$  opløftet i ottende, afrundet til nærmeste heltal, og herefter fratrukket 1. Tanken hertil er et ønske om en eksponentielt stigende maksimal længde fra højeste zoomniveau til laveste zoomniveau.

### 5.2.2 Interessegrad for rækker

Rækkernes interessegrad fastsættes naivt efter deres cellers summerede absolutte værdier i forhold til summen af alle værdier. Der ønskes en nær én til én korrespondance mellem hvor langt der zoomes ud og hvor mange celler der

fravælges. Således vælges de  $u$  procent mindst signifikante rækker ud til at være skjulte på et givent niveau, hvor  $u = \max(1 - k \cdot s^2, 0)$ .  $k$  er en konstant som påvirker startniveauet for hvornår rækker fravælges - det er ikke ønskeligt at fravælge rækker på zoomniveauer nær det maksimale zoomniveau da da brugeren skal have lov til at indstille sig på den igangværende skalering inden data sorteres fra.

## 6 Implementering af udvalgte metoder

Til implementering af løsningsforslagene fra afsnit 5, er benyttet Piccolo .NET 1.2 [8], en efterfølger til Jazz [9], præsenteret af B.B. Bederson et al. [7]. Piccolo er et værktøj til at lette udviklingen af applikationer som benytter zoombare grænseflader.

Den udvalgte løsning er, som nævnt tidligere, blevet testet sammen med en referenceløsning som kun benytter sig af geometrisk zoomning. Denne løsning er grundlaget for de resterende implementationer, hvorfor den primære virkemåde af begge grænseflader kun er beskrevet i afsnit 6.1 herunder.

### 6.1 Løsning med geometrisk zoomning

I figur 17 i afsnit A.1 på side 41 ses et skærmbillede af referenceløsningen. Indhold i dataceller er højrejusteret og indhold i titelceller er venstrejusteret. Endvidere er dataarket indlæst med titler som sikrer en hierarkisk opdeling af rækker - ganske som ville være tilfældet ved import i gængse programmer. Den valgte skrifttype er Arial i punkt 10 som benyttes som standard i Microsoft Excel 2003. Øverst er implementeret en værktøjslinje som giver adgang til regnemaskinen (se afsnit 6.3) samt lukning af applikationen (brugt da brugerne skifter mellem flere dataark i forsøgsfasen). Nedrest findes en statuslinje som i referenceløsningen blot angiver det pågældende zoomniveau.

Tekst er skrevet med sort på hvid baggrund og cellerne er indrammet med samme lysegrå kant som benyttes i Microsoft Excel 2003.

Navigering i grænsefladen foregår med musen. Der bevæges rundt i dataarket ved at gribe fat i det med musens venstre knap og derefter bevæge den i den ønskede retning. Zoomning foregår med musens venstre knap hvor bevægelse mod højre zoomer ud og bevægelse mod venstre zoomer ind. Disse navigationsmuligheder er standardimplementationer valgt af udviklerne af Piccolo.

For at sikre at brugerne altid kan se til hvilken række og hvilken kolonne en given celle hører er det valgt at gøre række- og kolonneoverskrifter statiske i forhold til applikationens sider - således vil f.eks. årstal ikke forsvinde uden for det synlige område selvom brugeren trækker dataarket op.

Endvidere er det valgt at lade celler og deres tilhørende række- og kolonneoverskrifter blive svagt markeret når musen holdes over. Dette er for at

	1990	1991	1992	1993	1994
Eksport af varer og tjenester	365.015	387.411	387.400	389.862	
- Eksport af tjenester	106.855	117.339	114.336	111.929	
- Eksport af varer	258.160	270.072	273.064	277.933	
Faste bruttoinvesteringer mv.	191.632	184.962	179.525	169.388	
Import af tjenester	93.246	92.549	92.944	98.460	
- Import af varer	229.968	240.806	234.183	224.801	
- Import af varer og tjenester	323.214	333.355	327.127	323.261	
Forbrug i produktionen	799.732	802.100	810.673	819.608	
- Forbrugsudgift	801.695	810.438	824.143	825.956	
- Privat forbrugsudgift	536.202	542.676	554.598	545.172	
- Offentlig individuel forbrugsudgift	174.297	175.525	180.207	186.914	
- Offentlig kollektiv forbrugsudgift	91.196	92.238	89.338	93.869	
- Offentlig forbrugsudgift	265.493	267.763	269.545	280.783	
Lagerændringer	(-511)	(-1.488)	(-693)	(-6.801)	
Produktion	1.682.697	1.693.367	1.714.153	1.724.156	1.8
- Produktion til eget brug	79.218	80.812	82.799	84.999	
- Markedsrettet produktion	1.340.312	1.348.109	1.365.452	1.362.331	1.4
- Anden ikke-markedsrettet produktion	263.166	264.445	265.902	276.826	
Produktskatter	24.995	25.374	25.965	26.068	
- Produktskatter minus produktsubsidier	151.653	156.703	159.768	150.595	
Produktsubsidier	126.658	131.329	133.803	124.527	

Figur 15: Skærbillede af grænseflade med semantisk zoomning

sikre at brugeren med lethed kan se hvortil en given celle hører til trods for at den muligvis er langt fra overskrifter.

Alle andre indstillinger og størrelser af elementer i grænsefladen er standard ved udvikling af grænseflader til Microsoft Windows XP.

Begge løsninger har en fastsat minimum zoomniveau til 75 % og maksimalt zoomniveau til 100 % på baggrund af læseligheden af teksten. Ved et zoomniveau på 75% af de oprindelige størrelser er teksten ved at være på grænsen til om det er lige til at læse eller ej. Endvidere giver zoomniveauer på over 100% ikke mening i forhold til overskueligheden af dataarket.

## 6.2 Løsning med semantisk zoomning

Skærbillede af løsningen med semantisk zoomning ses i figur 15. I forhold til referenceløsningen er der tilføjet to nye felter i statuslinjen. Her er en angivelse af en celles indhold samt antallet af celler som er indholdt heri<sup>8</sup> samt et felt som angiver datarepræsentationen på det givne zoomniveau.

Eftersom denne grænseflade lægger op til at rækker frasorteres er der endvidere implementeret mulighed for at brugeren kan vælge rækker som ikke må blive medtaget i denne frasortering. Udvalgelse og fravælgelse af sådanne rækker sker ved simpelt tryk med musens venstre knap på en vilkårlig celle i rækken. Rækken markeres da svagt for at indikere at den er valgt. Således

<sup>8</sup> dette er implementeret for at brugeren har mulighed for at se de reelle værdier i en given celle til trods for at denne celle er et resultat af en sammenfletning af flere celle

	1990 - 1992	1993	1994 - 1998	1999 - 2000	2001 - 2002	2003 - 2005
Eksport af varer og tjenester					4,5%	3,7%
- Eksport af varer					4,1%	3,7%
Faste bruttoinvesteringer mv.			-3%	6,3%	3,7%	3%
- Import af varer	0,8%				5,8%	5,3%
- Import af varer og tjenester	0,4%				6%	5,7%
Forbrug i produktionen					2,6%	2,4%
- Forbrugsudgift						1,6%
- Privat forbrugsudgift	1,1%					1,8%
- Offentlig forbrugsudgift						1,7%
Produktion						2,2%
- Måkedsmæssig produktion						-100%
- Anden ikke-måkedsmæssig produktion						-100%
Anvendelse i alt	0,7%				3,2%	2,7%

Figur 16: Skærbillede af dataark på laveste zoomniveau

ved brugeren at den på et senere tidspunkt kan fravælges for ikke at påvirke den endelige repræsentation i fald den ikke indgår i gruppen af de synlige rækker.

I figur 16 ses et af dataarkene fra brugerforsøgene når der er zoomet helt ud og ingen rækker er valgte.

Som det ses er der sket en betydelig reduktion i antallet af celler - fra 476 celler til blot 26 celler - altså en reduktion af celler hhv. areal på 94,4 % og 82,4 %. I appendiksets afsnit A.2 ses en række skærbilleder taget med samme dataark på forskellige zoomniveauer.

### 6.3 Regnemaskine

For at lette forsøgspersonernes svar på de konstruerede opgaver er det valgt at implementere en regnemaskine specifikt til forsøget. Regnemaskinens grænseflade ses i figur 18. Brugeren indtaster blot værdierne i de tilhørende felter og så beregnes resultatet løbende. Denne regnemaskine er valgt for ikke forurene resultaterne med brugernes individuelle beregning af sådanne tal i hånden eller på en eventuel lommeregner. Fejlraten ved fejlregning bør derfor være elimineret hvorfor fejl i besvarelserne burde kun kunne skyldes forkert indtastning, forståelsesproblemer eller forkert tolkning af dataarket.

## 7 Evaluering af implementerede metoder

Nærværende afsnit redegør for de metoder der er benyttet i forbindelse med evalueringen af det implementerede løsningsforslag til en applikation der benytter sig af semantisk zoomning til præsentation af statistisk materiale i tabelform.

### 7.1 Kvantificering af metodernes brugbarhed

For at kvantificere problemformuleringens *'i hvilket omfang'* er udvalgt en række målbare egenskaber ved interaktionen med grænsefladen. De udvalgte

egenskaber, taget fra usability-analysen [15], er effektivitet, fejlrate og tilfredshed. I forbindelse med forsøget er dette tre egenskaber som er relativt ligetil at måle. Ved løsning af opgaver kan effektiviteten af grænsefladen ses som direkte influerende på svartiden på hver opgave og fejlraten kan måles ved vurdering af besvarelsens korrekthed. Den subjektive del - tilfredshed - er der også ganske fine værktøjer til at måle med, mere herom i afsnit A.5. Forsøgspersonernes interaktionsmønstre er endvidere prøvet fastholdt.

## 7.2 Udvalgt sammenligningsgrundlag

Det er valgt at lave en dobbeltundersøgelse for at sammenligne den udviklede løsning med en referenceløsning. Hertil er brugt den tidligere nævnte løsning med geometrisk zoomning. I realiteten må andre applikationer bedre kunne vurderes som alternativer til den udviklede løsning - heriblandt regnearksprogrammer med mulighed for optegning af grafer over dataene. Det er dog valgt at sammenligne med en løsning uden mulighed for sådanne andre repræsentationsformer. Grunden hertil er et ønske om at sammenligne princippet om semantisk zoomning i forhold til den gængse geometrisk zoomning - det er altså ikke målet at vurdere løsningen imod metoder meget fjernt herfra. Den udviklede løsning må mest betragtes som et alternativ til den gængse regnearkspræsentation fremfor et alternativ til eksempelvis plotning i grafer.

## 7.3 Evalueringsform

Det udviklede eksperimentelle design evalueres som nævnt tidligere i forhold til de tre usabilityfaktorer effektivitet, fejlrate og tilfredshed. De første to er målt ved at stille forsøgspersonen en række opgaver i henholdsvis grænsefladen med den udviklede løsning og grænsefladen med referenceløsningen. Under løsningen af denne opgave tages da tid således at forsøgspersonernes svartider på de enkelte opgaver er kendte. Disse kan da bruges til at måle effektiviteten af grænsefladerne. De af forsøgspersonen nedskrevne svar vurderes da efterfølgende i forhold til en facitliste hvorved et udtryk for fejlraten kan findes. Endeligt gives forsøgspersonerne et spørgeskema der søger at klarlægge deres subjektive tilfredshed med grænsefladerne. Forsøgene var pr. person på en time - hertil omkring 15 minutters introduktion 35 minutters opgaveløsning og 10 minutters afslutning.

I de følgende afsnit vil de enkelte dele af evalueringen blive uddybet nærmere.

### 7.3.1 Introduktion

Forud for undersøgelserne blev deltagerne introduceret for de to applikationer samt formålet med undersøgelsen. For at gøre forsøgspersonerne bekendte med grænsefladerne forklaredes interaktionsmåden og forsøgspersonerne stilledes en række opvarmningsopgaver. Forsøgspersonen blev gjort op-

mærksom på at der var tale om en undersøgelse der søger at afklare hvorvidt princippet om semantisk zoomning er brugbart i den givne situation eller ej. Derudover fulgtes de generelle vejledninger omkring udførsel af forsøg angivet i [16].

Som introduktion til grænsefladerne stilledes forsøgspersonerne en række opvarmningsopgaver. Disse er konstrueret med mål for at give forsøgspersonen fortrolig med grænsefladerne og interaktionsmetoderne. De stillede opvarmningsopgaver var som følger.

1. Vælg 'Geometrisk zoomning - Introduktion' i opstartsvinduet<sup>9</sup>
  - (a) Find cellen i nederste højre hjørne og aflæs indholdet.
  - (b) Find øverste venstre celle og aflæs indholdet.
  - (c) Vha. regnemaskinen beregn den årlige gennemsnitlige vækstrate fra øverste højre celle og fire celler frem - aflæs resultatet
  - (d) Vha. regnemaskinen beregn den relative ændring fra øverste højre celle til cellen til venstre herfor - aflæs resultatet
2. Luk dataarket og regnemaskinen og vælg nu 'Semantisk zoomning - Introduktion' i opstartsvinduet
  - (a) Find cellen i nederste venstre hjørne - aflæs indholdet.
  - (b) Zoom helt ud og naviger derefter op til cellen i øverste højre hjørne.
  - (c) Zoom helt ind på cellen i øverste venstre hjørne og aflæs indholdet.
  - (d) Zoom helt ud på arket et vilkårligt sted, vælg tre tilfældig rækker ved at klikke på dem og zoom da helt ind igen.
  - (e) Zoom helt ud og naviger op til cellen i øverste venstre hjørne. Peg på cellen og aflæs dens indeholdte værdier i feltet under arket.
  - (f) Angiv den årlige gennemsnitlige vækstrate fra cellerne i række 4 fra søjle 7 til søjle 11
  - (g) Angiv den absolutte ændring af cellerne i række 9 fra søjle 2 til søjle 3

Under de opvarmende spørgsmål havde forsøgspersonen mulighed for at stille spørgsmål til opgaverne samt interaktionsmåden. Efter introduktionen gives en kort indledning til de følgende opgaver. Af valgte data til opvarmningsspørgsmålene er blot genereret et dataark med tilfældige tal bestående af 15 rækker og søjler.

---

<sup>9</sup>Opstartsvinduet var blot en maksimeret mappe med en række genveje til de forskellige grænseflader og dataark.

$I_s$	$O_1$	$I_g$	$O_1$
$I_s$	$O_2$	$I_g$	$O_1$
$I_s$	$O_{1A}$	$I_g$	$O_1$
$I_g$	$O_1$	$I_s$	$O_2$
$I_g$	$O_1$	$I_s$	$O_{1A}$
$I_g$	$O_2$	$I_s$	$O_1$
$I_g$	$O_{1A}$	$I_s$	$O_2$

Tabel 1: Rækkefølge af grænseflade og opgavesæt ved forsøg

### 7.3.2 Opgaver

I det følgende præsenteres de opgaver som forsøgspersonerne blev bedt om at løse. Opgaverne er inddelt i en række klasser for at lette analysen af resultaterne samt have mulighed for at klassificere hvilke typer opgaver den ene grænseflade støtter bedre end den anden. Forsøgspersonerne var under undersøgelse dog ikke opmærksom på denne opdeling. For at sikre relativt upartiske resultater udførtes forsøgene som angivet i tabel 1.

Hvor  $I_s$  er grænsefladen med semantisk zoomning,  $I_g$  er grænsefladen med geometrisk zoomning,  $O_1$  er opgavesæt 1,  $O_2$  er opgavesæt 2 og  $O_{1A}$  er opgavesæt 1A. Herved sikres bedst muligt at forhold som opgaveforskel og startgrænseflade ikke påvirker resultaterne.

Typen af opgaver der stilles kan inddeles i følgende klasser

1. Navigeringsopgaver
2. Trendanalyseopgaver
3. Opgaver med fokus på opdeling i delperioder

I det følgende er opstillet de spørgsmål som udgør de opgavesæt forsøgspersonerne gennemførte. I parentes efter hvert spørgsmål er endvidere angivet til hvilken gruppe i listen ovenover spørgsmålet hører.

**Opgavedataene** Af dataark til opgaverne er valgt de rækker fra nationalregnskabet som udgør den lange udgave af konto 1 - vare- og tjenestekonto [17] - se bilag A.3 for en gengivelse af dataarkene. Disse data er udvalgt for deres relevans i forhold til virkelige opgaver samtidig med at der her foreligger data tilbage til 1974, uden brud, således at to delperioder af en rimelig længde kan udvælges. Første del af opgavesættet benytter sig da af årene 1974 til 1989 og anden del benytter sig af årene 1990 til 2005. Begge delark består således af 22 rækker og 16 kolonner, foruden overskrifter, altså i alt 352 celler. Denne størrelse er valgt for at forsøgspersonerne har mulighed for at løse alle opgaverne inden for de 35 afsatte minutter.

### Opgavesæt 1

1. Angiv den absolutte ændring i 'Import af varer og tjenester' fra 1992 til 2002 (1)
2. Angiv 'Anvendelse i alt' for 2005 (1)
3. Angiv den gennemsnitlige årlige vækstrate over hele perioden for 'Eksport af varer' (2)
4. Angiv den gennemsnitlige årlige vækstrate fra 2001 til 2005 for 'Faste bruttoinvesteringer mv.' (2)
5. Angiv den relative ændring i forbrugsudgiften fra 1999 til 2000 (1)
6. Opdel 'Import af varer og tjenester' i relevante delperioder. (3)
7. Angiv 'Produktion til eget brug' i 2002 (1)
8. Angiv den relative ændring i 'Forbrug i produktionen' fra 1990 - 1992 (1)
9. Angiv den gennemsnitlige årlige vækstrate for 'Forbrug i produktionen' fra 1990 til 1994 (2)
10. Angiv den gennemsnitlige årlige vækstrate over hele perioden for 'Produktion' (2)

### Opgavesæt 2

1. Angiv den absolutte ændring i 'Offentlig kollektiv forbrugsudgift' fra 1975 til 1987 (1)
2. Angiv 'Tilgang i alt' for 1983 (1)
3. Angiv den gennemsnitlige årlige vækstrate over hele perioden for 'Produktskatter minus produktsubsidier' (2)
4. Angiv den gennemsnitlige årlige vækstrate fra 1980 til 1984 for 'Privat forbrugsudgift' (2)
5. Angiv den relative ændring i 'Forbrugsudgift' fra 1974 til 1975 (1)
6. Opdel 'Import af varer og tjenester' i relevante delperioder. (3)
7. Angiv 'Markedsmæssig produktion' i 1989 (1)
8. Angiv den relative ændring i 'Import af tjenester' fra 1980 - 1982 (1)
9. Angiv den gennemsnitlige årlige vækstrate for 'Eksport af varer' fra 1978 til 1981 (2)

10. Angiv den gennemsnitlige årlige vækstrate over hele perioden for 'Faste bruttoinvesteringer mv.' (2)

Opgavesæt 1A svarer til opgavesæt 1 blot med årstallene erstattet med årstallene fra opgavesæt 2. Der er altså i hvert opgavesæt fem navigeringsopgaver, fire opgaver med fokus på beskrivelse af trend og én opgave med fokus på opdeling i delperioder. Forsøgspersonerne blev gjort opmærksom på at de skulle løse opgaverne hurtigst muligt og uden at spørge undervejs. På opgavearket var afsat plads til nedskrivning af svar hvortil forsøgspersonen havde fået udleveret kuglepen. Hvert opgavesæt er planlagt til at tage omkring 15 minutter at løse. I afsnit A.4 på side 49 ses eksempler på de udleverede opgave- og svarark.

For at fastholde forsøgspersonernes interaktionsmønstre er der under tidstagningen endvidere også nedfældet hvorledes forsøgspersonerne zoomer. Således er der anført et plus ved zoomen ind og minus for zoomen ud. Dette er gjort for i analysen at kunne kvantificere hvor meget forsøgspersonerne gjorde brug af zoomfunktionaliteten.

### 7.3.3 Spørgeskema

For at kvantificere den subjektive opfattelse af applikationen gives efter opgaven et spørgeskema baseret på et udvalg fra QUIS [13] samt spørgsmål specifikt rettet mod den udviklede grænseflade. Inspiration er ligeledes hentet fra [14]. Alle spørgsmålene kunne vægtes fra nul til 9. Spørgsmålene var som følger

1. Hvordan var grænsefladen at bruge
  - (a) Forfærdelig - vidunderlig
  - (b) Svær - nem
  - (c) Frustrerende - tilfredsstillende
  - (d) Kedelig - sjov
  - (e) Ufleksibel - fleksibel
  - (f) Langsom - hurtig
2. Grænsefladen var svær at lære
  - (a) Helt enig - helt uenig
3. Hvordan opfattede du de stillede opgaver
  - (a) Meget udfordrende - meget nemme
  - (b) Aldrig ligetil at løse - altid ligetil at løse

4. Var dataarket...
  - (a) Svært at forstå - nemt at forstå
  - (b) Svært at skabe sig et overblik over - nemt at skabe sig et overblik over
5. Specifikke data var...
  - (a) Meget svære at finde - meget lette at finde

Et spørgeskema for både referenceløsningen samt den udviklede løsning blev udfyldt. Udover den subjektive del indeholdte spørgeskemaet også felter til udfyldelse af følgende informationer

1. Forsøgspersonens alder
2. Forsøgspersonens køn
3. Forsøgspersonens erfaring med at arbejde med statistisk materiale og regnearksapplikationer - vurderet på en skala fra 0 til 9
4. Angivelse af maksimalt tre fordele ved løsningen med semantisk zoomning i forhold til referenceløsningen.

I afsnit A.5 på side 50 ses eksempler på spørgeskemaerne.

#### **7.3.4 Efter forsøget**

Efter at alle opgaver var løst og spørgeskemaet udfyldt fik forsøgspersonen mulighed for, verbalt, at kommentere på det udviklede program samt stille eventuelle spørgsmål. Endeligt takkedes pænt af.

#### **7.3.5 Udstyr**

Alle opgaverne blev udført på en bærbar computer med tilsluttet standard dansk keyboard samt en optisk trådet mus. Computerens skærmopløsning var sat til 1024 x 768 pixels og applikationen optog godt 4/5 af skærmens areal for at der var plads til regnemaskinen ved siden af applikationen. Tidstagning foregik med stopur af forsøgslederen.

#### **7.3.6 Forsøgspersoner**

Som forsøgspersoner var hvervet en række studerende fra Roskilde Universitetscenter. Det drejer sig om studerende fra retninger som datalogi, kommunikation, økonomi, samfundsvidenskab, fysik og matematik. Alle forsøgspersonerne er irgang med den afsluttende del af bacheloruddannelsen eller i gang med kandidatuddannelsen og har derved alle et rimeligt grundlag for at kunne magte de stillede opgaver. Af forsøgspersonerne var der 3 kvinder og 4 mænd og gennemsnitsalderen var 23 år.

	Geometrisk		Semantisk		$p$
	$M$	$SD$	$M$	$SD$	
Opgavetype 1 - Navigering	52,1	37,1	69,0	46,6	0,257
Opgavetype 2 - Trendanalyse	59,2	25,2	76,5	56,7	0,321
Opgavetype 3 - Opdeling i delperioder	220,3	95,7	151,0	58,2	0,128

Tabel 2: Gennemsnitlig opgavegennemførelsestid, standardafvigelse herfra og  $p$  for de to grænseflader

## 8 Resultat af evaluering

I dette afsnit analyseres og diskuteres de data der er indsamlet ved metoderne i forrige afsnit. Endeligt afsluttes med en opsummering og kort diskussion af fremtidige muligheder.

### 8.1 Dataanalyse

I det følgende er der søgt at kaste lys på nogle af de hypoteser som umiddelbart kan konstrueres fra det indsamlede datagrundlag. For at understøtte, eller eventuelt forkaste hypoteserne er der brugt Students T-test med et ønsket signifikansniveau  $\alpha = 0,05$ , retningsuafhængighed samt antagelse om ens varians - med resultater angivet ved  $p$ <sup>10</sup>. Endvidere er værdier udtrykt som median ( $M$ ) og standardafvigelse ( $SD$ ).

#### 8.1.1 Effektivitet

Som beskrevet i afsnit 7.3.2 kan de løste opgaver opdeles i tre typer. I tabel 2 ses de gennemsnitlige gennemførelsestider, standardafvigelsen herfra og  $p$  fra t-testene for de to grænseflader.

Det ses at der umiddelbart ingen klar statistisk velfunderet forskel er mellem opgavegennemførelsestiden for de to grænseflader fordelt over de tre opgavetyper. Om end den gennemsnitlige udførelsestid for opgavetype 3 er 31 % hurtigere - er der dog kun svagt belæg for at dette er en generel tendens frem for tilfældighed. Ligeledes er der ingen signifikant grundlag for at støtte den hypotese som resultaterne for opgavetype 1 og 2 umiddelbart ligger op til - altså at grænsefladen med semantisk zoomning gennemsnitligt klarer disse 33 % hhv. 29 % ringere end grænsefladen med geometrisk zoomning. I tabel 7 på side 52 ses resultaterne fordelt på de enkelte opgaver.

Om end der ikke kan siges noget interessant med 95 % sikkerhed kan det dog stadig understøttes at opgavetype 3 må findes hurtigere at løse med

<sup>10</sup>For at en eventuel forskel skal være statistisk signifikant skal  $p \leq \alpha$ . Dette kan dog ikke betragtes som et bevis eller modbevis for hypotesen, men er blot et udtryk for om hypotesen er nogenlunde understøttet af datagrundlaget. Ved  $\alpha = 0,05$  og  $p \leq \alpha$  kan det altså ses at dataene er gode nok til at kunne understøtte en konklusion med 95 % sikkerhed.

	Geometrisk		Semantisk		$p$
	$M$	$SD$	$M$	$SD$	
Opgavetype 1 - Navigering	1,0	0,2	1,0	0,2	1,000
Opgavetype 2 - Trendanalyse	1,0	0,2	0,8	0,4	0,190
Opgavetype 3 - Opdeling i delperioder	0,1	0,4	0,7	0,5	0,031

Tabel 3: Gennemsnitlig fejlrate, standardafvigelse herfra og  $p$  for de to grænseflader

grænsefladen der benytter sig af semantisk zoomning.

### 8.1.2 Fejlrate

Til vurdering af de løste opgaver er der givet point pr. opgave pr. besvarelse - en korrekt opgave gives ét point mens en forkert opgave gives nul point. Om end dette må findes tilstrækkeligt i opgavetype 1 og 2 er det mere problematisk mht. opgavetype 3 - opdeling i relevante delperioder. Det er dog valgt at bruge samme konvention dog med nærmere granskning af de enkelte opgaver. Således gives en opgave af type 3 point i fald den ikke er direkte forkert - altså med arbitrære opdelinger - samt hvis den ikke overser relevante tendenser - altså beskriver ændringer i udviklingstendensen. I tabel 8 på side 53 ses de konstaterede fejl, tabel 9 på side 53 viser fejlraten for de enkelte spørgsmål og i tabel 3 ses fejlratene opdelt i grupper.

Begge grænseflader resulterede i samme antal korrekte svar - 62 ud af 70 (80,5 %) med begge grænseflader havende samme antal korrekte svar af opgavetype 1, grænsefladen med geometrisk zoomning flere korrekte af opgavetype 2 - dog intet grundigt statistisk belæg for at støtte en hypotese herom - samt grænsefladen med semantisk zoomning med flere rigtige svar i opgavetype 3. Som det ses i tabel 3 er der ganske godt belæg for at grænsefladen med semantisk zoomning giver flere korrekte svar med  $p = 0,031$ . Reelt så er der tale om en fejlrate på 86 % imod 29 % med semantisk zoomning. Der ses dog også en rimelig tendens imod at opgaver af type 2 tilsyneladende løses ringere med semantisk zoomning.

### 8.1.3 Subjektiv undersøgelse

I figur 4 ses forsøgspersonernes gennemsnitlige vurdering, fra den subjektive undersøgelse, af de to løsninger samt standardafvigelsen fra gennemsnittet. Til trods for at der ses variationer i de enkelte spørgsmål er det samlede gennemsnit ikke betydeligt forskelligt.

Vurderingerne summeret pr. person (se bilag A.8 på side 54) giver heller ingen tegn på statistisk signifikante forskelle,  $p = 0,884$ . Tabel 5 på den følgende side viser  $p$  for de enkelte spørgsmål.

I fald sikkerhed på 90% er acceptabelt vil det dog kunne argumenteres

	Geometrisk	Semantisk
<b>1. - 5. Hvordan var grænsefladen at bruge</b>		
Forfærdelig - Vidunderlig	4,71 (1,81)	5,43 (1,27)
Svær - Nem	7,29 (2,21)	5,29 (1,89)
Frustrerende - Tilfredsstillende	4,14 (2,27)	4,14 (2,19)
Ufleksibel - Fleksibel	3,86 (3,44)	6,00 (2,31)
Langsom - Hurtig	5,43 (2,88)	6,00 (1,53)
<b>6. Grænsefladen var svær at lære</b>		
Helt enig - Helt uenig	7,00 (2,38)	5,71 (2,06)
<b>7. - 8. Hvordan opfattede du de stillede opgaver</b>		
Aldrig ligetil at løse - Altid ligetil at løse	7,14 (1,57)	6,29 (1,98)
Meget udfordrende - Meget nemme	6,00 (2,16)	6,57 (1,99)
<b>9. - 10. Var dataarket...</b>		
Svært at forstå - Nemt at forstå	7,29 (1,70)	6,57 (1,40)
Svært at danne sig et overblik over - Nemt at danne sig et overblik over	4,57 (3,26)	7,14 (1,21)
<b>11. Specifikke data var...</b>		
Meget svære at finde - Meget nemme at finde	6,57 (3,36)	6,00 (1,73)
<i>Summerede vurderinger</i>		
	5,82 (2,69)	5,92 (1,86)

Tabel 4: Gennemsnitlig tilfredshedsvurdering (standardafvigelse) af de to grænseflader

Spørgsmål	1	2	3	4	5	6
<i>p</i> -værdi	0,408	0,094	1,000	0,196	0,651	0,301
Spørgsmål	7	8	9	10	11	
<i>p</i> -værdi	0,387	0,616	0,408	0,075	0,696	

Tabel 5: Subjektiv vurdering af grænsefladerne, *p*-værdier

Svartype	Hyppighed
'Grænsefladen giver større overblik'	5
'Nemmere løsninger til opgaver som kræver beskrivelse af udviklingstendens'	3
Ovenstående plus 'i fald at applikationens periodeopdeling er sammenfaldende med forespørgslen i opgaven'	2
'Sparer beregninger'	2

Tabel 6: Grupper af fordele ved grænseflade med semantisk zoomning

at spørgsmål 2 (Svær - nem) og spørgsmål 10 (Svært at danne sig et overblik over - Nemt at danne sig et overblik over) med hhv.  $p = 0,094$  og  $p = 0,074$  understøtter resultaterne om at forsøgspersonerne gennemsnitligt har vægtet spørgsmålene med hhv. 7,29 i forhold til 5,29 og 4,57 i forhold til 7,14 - angivet med geometrisk zoomning først.

#### 8.1.4 Præference

Resultaterne fra spørgeskemaerne viste at 3 foretrak grænsefladen med geometrisk zoomning hvorimod 4 foretrak grænsefladen med semantisk zoomning - altså 43 % i forhold til 57 %.

#### 8.1.5 Erfaringsniveau

Ved angivelsen af forsøgspersonernes erfaringsniveau på spørgeskemaet kan det ses at forsøgspersonerne gennemsnitligt vurderer sig selv til 5,83 på skalaen fra 0 til 9, med standardafvigelse på 2,79.

#### 8.1.6 Angivne fordele

Spørgeskemaet gav også mulighed for nedskrivning af fordele ved grænsefladen med semantisk zoomning i forhold til grænsefladen med geometrisk zoomning. Herudfra de grupperinger af svar som ses i tabel 6.

Derudover blev der på et ark svaret

*“Det er svært at angive nogle fordele ud fra kortvarig brug. Dette må nødvendigvis kræve at jeg beskæftiger mig mere med det. [...]”*

## 8.2 Resultat

Som de foregående afsnit viser er der ikke mange direkte klare forskelle i de målte egenskaber ved de to grænseflader. Nævnværdigt er dog resultaterne fra opgavetype 3 - opdeling i delperioder. Som det ses i afsnit 8.1.1 på side 35 løses denne opgavetype i forsøgene gennemsnitligt 31% hurtigere i

grænsefladen med semantisk zoomning. Om end dette ikke er klart understøttet af de indsamlede data er der dog med  $p = 0,128$  en indikator om at dette muligvis ville kunne understøttes ved nærmere studier. Fejlraten for opgavetype 3 ser dog derimod en stærk statistisk signifikant forskel ved brug af den udviklede grænseflade. Her findes 71 % rigtige imod blot 14 % rigtige - en forskel der understøttes med  $p = 0,031$ .

Under undersøgelserne viste det sig at forsøgspersonerne søgte at gøre brug af mulighederne i semantisk zoomning. Det blev observeret at forsøgspersonerne ved løsning af opgaver i grænsefladen med semantisk zoomning brugte tid på at eksperimentere med dens repræsentationsændringer for at se om det svar de søgte allerede var givet for dem. Dette resulterede dog oftest i at forsøgspersonerne zoomede helt ind til repræsentationsniveauet med den almindelige cellegengivelse hvorved beregningerne kunne foretages manuelt. Denne brug af grænsefladen kan derfor muligvis være en årsag til den øgede svartid ved løsning af opgaverne af type 1. Dog vil det sandsynligvis ikke kunne bortforklare hele forskellen eftersom det også observeredes at forsøgspersonerne gjorde anderledes brug af muligheden for at markere rækker, end forventet. Det kunne ses at forsøgspersonerne, i grænsefladen med semantisk zoomning, valgte at markere rækker for nemmere at kunne finde den ønskede celle ved indtastning i regnemaskinen. Denne utiltænkte brug kan have haft uheldig indvirkning på de indsamlede resultater. Et problem som muligvis kunne være undgået ved en mere diskret markering af valgte rækker - eventuelt blot et lille ikon i rækkeoverskriften.

I forhold til den subjektive del var der et lille flertal der fortrak grænsefladen med semantisk zoomning - fire imod tre. Om end den gennemsnitlige vurdering via spørgeskemaerne peger lidt i denne retning er der dog ikke umiddelbart noget solidt statistisk belæg for en konklusion herfor. At referencerne vægter mest mod den udviklede grænseflade kan være simpel bias i forhold til forsøgspersonernes ubevidste ønske om at tilfredsstille forsøgslederen. Der må dog trods dette være belæg for at forsøgspersonerne finder en klar fordel i grænsefladens måde at præsentere dataene på ved yderste zoomniveau. Som afsnit 8.1.3 på side 36 påpeger er der en rimelig vægt imod at forsøgspersonerne finder den udviklede grænseflade mere hensigtsmæssig til at danne overbliksbilleder. Dette understøttes også ved forsøgspersonernes angivelse af fordele jf. afsnit 8.1.6 på forrige side. Disse subjektive udsagn understøttes ligeledes af de tidligere nævnte resultater fra løsninger af opgaver af type 3. Det viser sig dog også at forsøgspersonerne umiddelbart finder den udviklede grænseflade sværere at benytte hvilket umiddelbart også kan ses i svartider og fejlratere.

### 8.2.1 Videre undersøgelser

For at få et bedre billede af mulighederne ved semantisk zoomning på statistiske dataark er der behov for yderligere undersøgelser. Der er et behov for

at bedre statistisk datagrundlag for at kunne belyse de fundne resultater mere klart. De umiddelbart dårlige resultater i forhold til opgavetype 2 kan muligvis skyldes at forsøgspersonerne kun fik begrænset med tid til at arbejde med grænsefladen - noget som også påpeges af en forsøgsperson jf. afsnit 8.1.6 på side 38. Det vil være interessant at se om brugerne, med mere erfaring med grænsefladen - og derved bedre mulighed for at tolke de data der vises - vil kunne drage bedre nytte af den i denne type af opgaver.

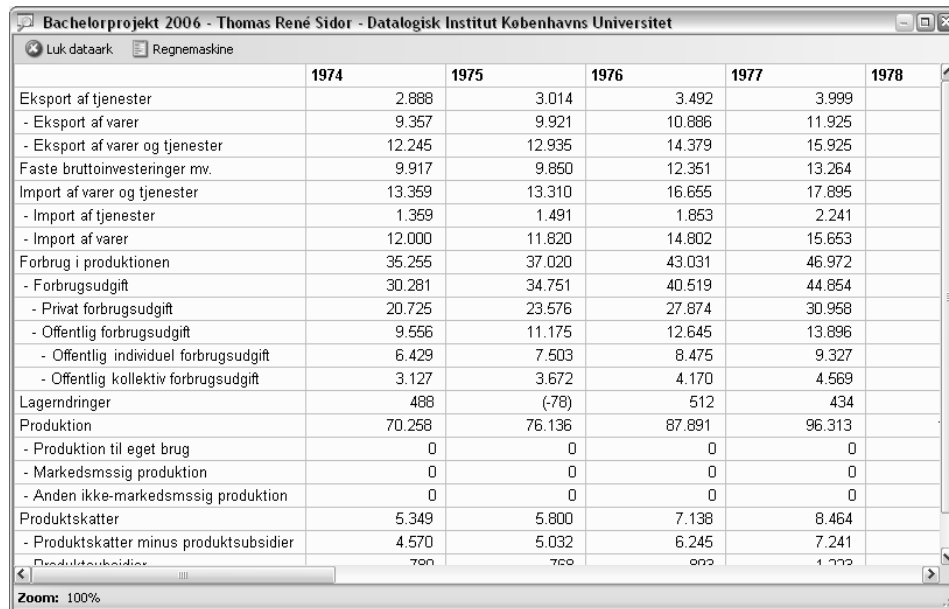
Derudover vil der være et behov for et bedre statistisk grundlag for at understøtte de resultater fundet i forbindelse med opgavetype 3. Blot at have én opgave af denne type i undersøgelsen er utilstrækkeligt i forhold til at kunne give nogle klare svar.

Endvidere vil der være et behov for en eventuel videreudvikling af de metoder der er gjort brug af. Det kan tænkes at de implementerede metoder jf. afsnit 5 på side 21 kommer til kort eller er direkte uhensigtsmæssige på nogle områder - noget som ikke kan ses direkte af resultaterne i nærværende opgave.

Det blev observeret at forsøgspersonerne kun meget lidt reelt brugte de mellemliggende zoomniveauer og -repræsentationer. Dette kunne muligvis pege i retning af at selve tanken om en flydende overgang fra referencerepræsentation til en beregnet repræsentation er mindre brugbar. I så fald kunne det tænkes at et par foruddefinerede niveauer ville være nok hvor brugeren så har mulighed for at aktivere eller deaktivere denne funktionalitet.

## A Bilag

### A.1 Skærbilleder



	1974	1975	1976	1977	1978
Eksport af tjenester	2.888	3.014	3.492	3.999	
- Eksport af varer	9.357	9.921	10.886	11.925	
- Eksport af varer og tjenester	12.245	12.935	14.379	15.925	
Faste bruttoinvesteringer mv.	9.917	9.850	12.351	13.264	
Import af varer og tjenester	13.359	13.310	16.655	17.895	
- Import af tjenester	1.359	1.491	1.853	2.241	
- Import af varer	12.000	11.820	14.802	15.653	
Forbrug i produktionen	35.255	37.020	43.031	46.972	
- Forbrugsudgift	30.281	34.751	40.519	44.854	
- Privat forbrugsudgift	20.725	23.576	27.874	30.958	
- Offentlig forbrugsudgift	9.556	11.175	12.645	13.896	
- Offentlig individuel forbrugsudgift	6.429	7.503	8.475	9.327	
- Offentlig kollektiv forbrugsudgift	3.127	3.672	4.170	4.569	
Lagerdringer	488	(-78)	512	434	
Produktion	70.258	76.136	87.891	96.313	
- Produktion til eget brug	0	0	0	0	
- Markedsmsig produktion	0	0	0	0	
- Anden ikke-markedsmsig produktion	0	0	0	0	
Produktskatter	5.349	5.800	7.138	8.464	
- Produktskatter minus produktsubsidier	4.570	5.032	6.245	7.241	
Produktsubsidier	700	700	800	1.000	

Figur 17: Referenceløsningens grænseflade

**Regnemaskine**

**Gennemsnitlig årlig vækstrate**

Startværdi

Slutværdi

Perioder

Resultat

**Relativ ændring**

Startværdi

Slutværdi

Resultat

**Absolut ændring**

Startværdi

Slutværdi

Resultat

Figur 18: Regnemaskinens grænseflade

## A.2 Skærbilleder fra grænseflade med semantisk zoomning

Bachelorprojekt 2006 - Thomas René Sidor - Datalogisk Institut Københavns Universitet

Luk dataark | Regnemaskine

	1990	1991	1992	1993	1994
Eksport af varer og tjenester	365.015	367.411	367.400	389.862	
- Eksport af tjenester	106.855	117.339	114.336	111.929	
- Eksport af varer	258.160	270.072	273.064	277.933	
Faste bruttoinvesteringer mv.	191.632	184.962	179.525	169.388	
Import af tjenester	93.246	92.549	92.944	96.460	
- Import af varer	229.968	240.806	234.183	224.801	
- Import af varer og tjenester	323.214	333.355	327.127	323.261	
Forbrug i produktionen	799.732	802.100	810.673	819.608	
- Forbrugsudgift	801.695	810.438	824.143	825.956	
- Privat forbrugsudgift	536.202	542.676	554.598	545.172	
- Offentlig individuel forbrugsudgift	174.297	175.525	180.207	186.914	
- Offentlig kollektiv forbrugsudgift	91.196	92.238	89.338	93.869	
- Offentlig forbrugsudgift	265.493	267.763	269.545	280.783	
Lagendringer	(-511)	(-1.488)	(-693)	(-6.801)	
Produktion	1.682.697	1.693.367	1.714.153	1.724.156	1.6
- Produktion til eget brug	79.218	80.812	82.799	84.999	
- Markedsmsig produktion	1.340.312	1.348.109	1.365.452	1.362.331	1.4
- Anden ikke-markedsmsig produktion	263.166	264.445	265.902	276.826	
Produktskatter	24.995	25.374	25.965	26.068	
- Produktskatter minus produktsubsidier	151.653	156.703	159.768	150.595	

Zoom: 100% Celler: 92.238 Enheder og informationer: Oprindelige værdier (mio. kr)

Figur 19: Skaleringsseksempel 1, Semantisk zoomning

Bachelorprojekt 2006 - Thomas René Sidor - Datalogisk Institut Københavns Universitet

Luk dataark Regnemaskine

	1990 - 1991	1992	1993	1994	1995
Eksport af varer og tjenester		365 - 387			390 - 427
- Eksport af tjenester		107 - 114			112 - 116
- Eksport af varer		258 - 273			278 - 311
Faste bruttoinvesteringer mv.		192 - 180			169 - 202
Import af tjenester	93 - 93			93 - 107	
- Import af varer		230 - 234			225 - 271
- Import af varer og tjenester		323 - 327			323 - 379
Forbrug i produktionen		800 - 811			820 - 883
- Forbrugsudgift		802 - 824			826 - 875
- Privat forbrugsudgift		536 - 555			545 - 583
- Offentlig individuel forbrugsudgift		174 - 180			187 - 196
- Offentlig kollektiv forbrugsudgift	91 - 92			89 - 97	
- Offentlig forbrugsudgift		265 - 270			281 - 292
Produktion		1.683 - 1.714			1.724 - 1.855
- Produktion til eget brug		79 - 93			85 - 89
- Markedsmsig produktion		1.340 - 1.365			1.362 - 1.478
- Anden ikke-markedsmsig produktion		263 - 286			277 - 288
- Produktskatter minus produktsubsidier		152 - 160			151 - 167
Anvendelse i alt		2.158 - 2.201			2.198 - 2.401

Zoom: 92% Celler: (3 celler) 169.388, 181.393, 201.836 Enheder og informationer: Oprindelige værdier (mio. kr) og intervaller (mia. kr)

Figur 20: Skaleringseksempel 2, Semantisk zoomning

Bachelorprojekt 2006 - Thomas René Sidor - Datalogisk Institut Københavns Universitet

Luk dataark Regnemaskine

	1990 - 1992	1993	1994 - 1998	1998	1999	2000
Eksport af varer og tjenester			3%, M: 416			
- Eksport af tjenester	2,3%, M: 116		3,3%, M: 119			
- Eksport af varer			3,3%, M: 298			
Faste bruttoinvesteringer mv.		-3%, M: 178		6,3%, M: 222		3,7%, M: 263
- Import af varer	0,6%, M: 237					6,2%, M: 303
- Import af varer og tjenester	0,4%, M: 330					6,2%, M: 427
Forbrug i produktionen			2%, M: 860			
- Forbrugsudgift			1,8%, M: 858			
- Privat forbrugsudgift	1,1%, M: 549					1,6%, M: 603
- Offentlig individuel forbrugsudgift			2%, M: 191			
- Offentlig forbrugsudgift			1,7%, M: 286			
Produktion			1,9%, M: 1.807			
- Markedsmsig produktion			1,9%, M: 1.438			
- Anden ikke-markedsmsig produktion			1,6%, M: 282			
- Produktskatter minus produktsubsidier	1,8%, M: 158				3,1%, M: 176	
Anvendelse i alt	0,7%, M: 2.192					3,4%, M: 2.571

Zoom: 83% Celler: (5 celler) 181.393, 201.836, 210.626, 229.398, 246.489 Enheder og informationer: Gennemsnitlig årlig vækstrate, median (mia. kr)

Figur 21: Skaleringseksempel 3, Semantisk zoomning

Bachelorprojekt 2006 - Thomas René Sidor - Datalogisk Institut Københavns Universitet

Luk dataark | Regnemaskine

	1990 - 1992	1993	1994 - 1998	1999 - 2000	2001 - 2002	2003 - 2005
Ekseport af varer og tjenester					4,5%	3,7%
- Ekseport af varer					4,1%	3,7%
Faste bruttoinvesteringer mv.			-3%	0,3%	3,7%	3%
- Import af varer	0,6%				5,8%	5,3%
- Import af varer og tjenester	0,4%				6%	5,7%
Førlug i produktionen					2,6%	2,4%
- Førlugsudgift						1,6%
- Privat forbrugsudgift		1,1%				1,8%
- Offentlig forbrugsudgift						1,7%
Produktion						2,2%
- Måkedsmæssig produktion						-100%
- Anden ikke-måkedsmæssig produktion						-100%
Anvendelse i alt	0,7%				3,2%	2,7%

Zoom: 75% Celler: (13 celler) 799.732, 802.100, 810.673, 819.608, 861.328, 883.405, 911.785, Enheder og informationer: Gennemsnitlig årlig vækstrate

Figur 22: Skaleringseksempel 4, Semantisk zoomning

### A.3 Dataark til brugertests

	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
Eksport af tjenester	2888	3014	3492	3999	4221	4806	5255	6888	7781	7856
- Eksport af varer	9357	9921	10886	11925	12846	15282	18992	22777	25861	29317
- Eksport af varer og tjenester	12245	12935	14379	15925	17066	20089	24248	29665	33643	37173
Faste bruttoinvesteringer mv.	9917	9850	13264	13264	14507	15636	15242	13949	16602	18713
Import af varer og tjenester	13359	13310	16655	17895	18346	22125	25100	29095	33309	35243
- Import af tjenester	1359	1491	1853	2241	2394	2892	3355	4339	5506	5561
- Import af varer	12000	11820	14802	15653	15953	19233	21745	24756	27803	29682
Forbrug i produktionen	35255	37020	43031	46972	51218	57741	65194	72840	81458	87748
- Forbrugsudgift	30281	34751	40519	44854	49769	55752	60976	67965	76924	83229
- Privat forbrugsudgift	20725	23576	27874	30958	33940	37813	40430	44571	49899	54178
- Offentlig forbrugsudgift	9556	11175	12645	13896	15829	17939	20545	23394	27025	29051
- Offentlig individuel forbrugsudgift	6429	7503	8475	9327	10619	12030	13774	15667	18097	19449
- Offentlig kollektiv forbrugsudgift	3127	3672	4170	4569	5209	5909	6771	7727	8928	9602
Lagerændringer	488	-78	512	434	-103	467	-77	-10	352	98
Produktion	70258	76136	87891	96313	105573	117440	129629	143355	162554	176751
- Produktion til eget brug	0	0	0	0	0	0	0	0	0	0
- Produktion til salg	0	0	0	0	0	0	0	0	0	0
- Markedsmæssig produktion	0	0	0	0	0	0	0	0	0	0
- Anden ikke-markedsmæssig produktion	0	0	0	0	0	0	0	0	0	0
Produktkatter	5349	5800	7138	8464	10001	11591	12290	13340	14711	16620
- Produktkatter minus produktsubsidier	4570	5032	6245	7241	8538	10120	10854	11959	13116	14968
- Produktsubsidier	780	768	893	1223	1464	1471	1436	1381	1595	1652
Anvendelse i alt	88186	94478	110791	121449	132457	149685	165583	184409	208979	226961

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
Eksport af tjenester	8383	9282	8545	9125	11329	12283	365015	387411	387400	389862
- Eksport af varer	33069	35812	34209	34802	37104	41639	106855	117339	114336	111929
- Eksport af varer og tjenester	41451	45094	42754	43927	48433	53921	258160	270072	273064	277933
Faste bruttoinvesteringer mv.	21804	25910	30418	31148	30195	31804	191632	184962	179525	169388
Import af varer og tjenester	40177	44788	43512	41440	44243	49162	93246	92549	92944	98460
- Import af tjenester	5921	6617	6667	6947	9001	9886	229968	240806	234183	224801
- Import af varer	34256	38170	36845	34493	35243	39277	323214	333355	327127	323261
Forbrug i produktionen	99687	109436	113252	115226	120208	128362	799732	802100	810673	819608
- Forbrugsudgift	89542	96509	102993	107870	111600	116519	801695	810438	824143	825956
- Privat forbrugsudgift	59227	64387	70002	71596	73267	76647	536202	542676	554598	545172
- Offentlig forbrugsudgift	30315	32122	32991	36274	38333	39872	174297	175525	180207	186914
- Offentlig individuel forbrugsudgift	20288	21505	21920	24323	25164	26181	91196	92238	89338	93869
- Offentlig kollektiv forbrugsudgift	10027	10618	11071	11951	13169	13691	265493	267763	269545	280783
Lagerændringer	1547	1274	1258	-901	-111	621	-511	-1488	-693	-6801
Produktion	197147	214815	225554	234209	243757	259679	1682697	1693367	1714153	1724156
- Produktion til eget brug	0	0	0	0	10672	11484	79218	80812	82799	84999
- Markedsmæssig produktion	0	0	0	0	195296	208818	1340312	1348109	1365452	1362331
- Anden ikke-markedsmæssig produktion	0	0	0	0	37788	39377	263166	264445	265902	276826
Produktskatter	18571	20298	23686	24106	24876	24896	24995	25374	25965	26068
- Produktskatter minus produktsubsidier	16707	18620	21608	21621	22325	22386	151653	156703	159768	150595
- Produktsubsidier	1864	1677	2078	2484	2552	2510	2573	2680	2804	3582
Anvendelse i alt	254030	278224	290674	297270	310324	331227	2157564	2183424	2201048	2198012

	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Eksport af tjenester	419391	426808	442440	460834	476637	534869	602351	621238	649935	642737
- Eksport af varer	114441	116113	127266	127138	133759	161939	196802	205733	213487	214260
- Eksport af varer og tjenester	304950	310695	315174	333696	342877	372930	405549	415505	436448	428477
Faste bruttoinvesteringer mv.	181393	201836	210626	229398	246489	244318	262880	259308	260014	265250
Import af varer og tjenester	107174	107887	112611	118891	129278	134784	159863	164130	181155	183984
- Import af tjenester	254049	270895	274490	302129	324694	330705	364390	370039	395894	383440
- Import af varer	361223	378782	387101	421020	453972	465489	524253	534168	577048	567425
Forbrug i produktionen	861328	883405	911785	934236	967942	1010095	1058897	1100382	1122859	1120775
- Forbrugsudgift	862723	875231	895095	912799	934882	937889	941781	949418	966857	977599
- Privat forbrugsudgift	576424	583261	593506	609524	621224	617668	616682	617185	627102	637057
- Offentlig forbrugsudgift	189624	196251	203450	204567	211963	217372	222219	227739	235282	237616
- Offentlig individuel forbrugsudgift	96675	95720	98139	98708	101695	102850	102880	104494	104473	102925
- Offentlig kollektiv forbrugsudgift	286299	291970	301589	303275	313658	320221	325099	332233	339755	340541
Lagerændringer	4065	13723	5897	16083	15171	1225	11204	7289	10948	2022
Produktion	1802708	1854942	1904543	1952006	2002164	2077011	2169795	2219467	2246496	2251728
- Produktion til eget brug	85830	88779	91426	94413	95897	100519	104162	103603	106358	0
- Markedsført produktion	1433993	1478425	1515933	1559094	1597947	1661925	1746792	1790828	1809203	0
- Anden ikke-markedsført produktion	282885	287737	297184	298499	308321	314566	318841	325035	330935	0
Produktskatter	28823	30066	32168	33951	36333	37619	37766	38815	40124	40554
- Produktskatter minus produktsubsidier	164969	167280	174198	180323	184986	185895	183065	184000	187068	189230
- Produktsubsidier	3576	3602	3464	3354	3375	3578	3475	3522	3431	3711
Anvendelse i alt	2328900	2401004	2465843	2553349	2641122	2728396	2877113	2937635	3010612	3008383

	2004	2005
Eksport af tjenester	659597	717050
- Eksport af varer	212769	239594
- Eksport af varer og tjenester	446828	477456
Faste bruttoinvesteringer mv.	276555	301130
Import af varer og tjenester	190271	223209
- Import af tjenester	413493	447752
- Import af varer	603763	670960
Forbrug i produktionen	1147367	1204170
- Forbrugsudgift	1006925	1035362
- Privat forbrugsudgift	660719	684936
- Offentlig forbrugsudgift	241927	249118
- Offentlig individuel forbrugsudgift	104278	101308
- Offentlig kollektiv forbrugsudgift	346205	350426
Lagerændringer	3920	5737
Produktion	2290463	2380782
- Produktion til eget brug	0	0
- Markedsmæssig produktion	0	0
- Anden ikke-markedsmæssig produktion	0	0
Produktskatter	42699	43012
- Produktskatter minus produktsubsidier	200137	211706
- Produktsubsidier	3773	3842
Anvendelse i alt	3094363	3263448

## A.4 Eksempel på svarark til spørgsmål

**Opgaver til løsning med semantisk zoomning**  
*Benyt venligst genvejen 'Semantisk Zoomning – Dataark 2' i opstartsvinduet*

**1. Angiv den absolutte ændring i 'Import af varer og tjenester' fra 1992 til 2002**  
Svar: \_\_\_\_\_

**2. Angiv 'Anvendelse i alt' for 2005**  
Svar: \_\_\_\_\_

**3. Angiv den gennemsnitlige årlige vækstrate over hele perioden for 'Eksport af varer'**  
Svar: \_\_\_\_\_

**4. Angiv den gennemsnitlige årlige vækstrate fra 2001 - 2005 for 'Faste bruttoinvesteringer mv.'**  
Svar: \_\_\_\_\_

**5. Angiv den relative ændring i 'Forbrugsudgift' fra 1999 til 2000**  
Svar: \_\_\_\_\_

**6. Opdel 'Import af varer og tjenester' i relevante delperioder.**  
Svar: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Figur 23: Eksempelside af opgaveark med svarplads

## A.5 Spørgeskema

**Grænseflade med semantisk zoomning**

**Hvordan var grænsefladen at bruge**

	0	1	2	3	4	5	6	7	8	9	
Forfærdelig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Vidunderlig
Svær	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Nem
Frustrerende	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Tilfredsstillende
Ufleksibel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Fleksibel
Langsom	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Hurtig

**Grænsefladen var svær at lære**

	0	1	2	3	4	5	6	7	8	9	
Helt enig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Helt uenig

**Hvordan opfattede du de stillede opgaver**

	0	1	2	3	4	5	6	7	8	9	
Aldrig ligetil at løse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Altid ligetil at løse
Meget udfordrende	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Meget nemme

**Var dataarket...**

	0	1	2	3	4	5	6	7	8	9	
Svært at forstå	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Nemt at forstå
Svært at danne sig et overblik over	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Nemt at danne sig et overblik over

**Specifikke data var...**

	0	1	2	3	4	5	6	7	8	9	
Meget svære at finde	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Meget nemme at finde

Side 2 af 3

Figur 24: Spørgeskema til subjektiv undersøgelse, side 1 og 2

**Sidste side**

Alder: \_\_\_\_\_

Køn: Mand  Kvinde

Hvilken af de to grænseflader fortrækker du: Semantisk zoomning   
Geometrisk zoomning

Angiv venligst hvor meget erfaring du har med at arbejde med statistisk materiale og regnearksapplikationer:

Meget erfaring    0   1   2   3   4   5   6   7   8   9    Ingen erfaring

Angiv op til tre fordele ved at bruge løsningen med semantisk zoomning frem for referenceløsningen:

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

Figur 25: Spørgeskema til subjektiv undersøgelse, side 3

## A.6 Datagrundlag for effektivitet

	Geometrisk		Semantisk		<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Opgave 1	63,0	19,2	102,7	51,5	0,080
Opgave 2	69,6	76,1	30,1	6,6	0,197
Opgave 3	66,4	26,0	114,7	85,7	0,179
Opgave 4	65,7	32,2	65,4	30,0	0,987
Opgave 5	55,7	14,9	70,3	31,3	0,288
Opgave 6	220,3	95,7	151,0	58,2	0,128
Opgave 7	29,0	4,4	62,4	48,3	0,0093
Opgave 8	43,1	13,4	79,6	55,1	0,115
Opgave 9	50,4	9,8	84,7	45,5	0,75
Opgave 10	54,1	28,4	41,3	27,7	0,408

Tabel 7: Gennemsnitlig opgavegennemførelstid, standardafvigelse herfra og *p* for de to grænseflader fordelt på de enkelte opgaver

## A.7 Datagrundlag til fejlrater

Forsøgsperson	1	2	3	4	5	6	7
<b>Semantisk zoomning</b>							
Opgave 1	1	1	1	1	1	1	1
Opgave 2	1	1	1	1	1	1	1
Opgave 3	1	1	0	0	1	1	0
Opgave 4	0	1	1	1	1	1	1
Opgave 5	1	1	1	1	1	1	1
Opgave 6	0	1	1	0	1	1	1
Opgave 7	1	1	1	1	1	1	1
Opgave 8	1	1	1	1	0	1	1
Opgave 9	1	1	1	1	1	1	1
Opgave 10	0	1	1	1	1	1	1
<b>Geometrisk zoomning</b>							
Opgave 1	1	1	1	1	1	1	1
Opgave 2	1	1	1	1	0	1	1
Opgave 3	1	1	1	1	1	1	1
Opgave 4	1	1	1	1	1	1	1
Opgave 5	1	1	1	1	1	1	1
Opgave 6	0	0	1	0	0	0	0
Opgave 7	1	1	1	1	1	1	1
Opgave 8	1	1	1	1	1	1	1
Opgave 9	1	1	1	1	0	1	1
Opgave 10	1	1	1	1	1	1	1

Tabel 8: Konstaterede fejl ved opgaveløsning

	<b>Geometrisk</b>		<b>Semantisk</b>	
	Antal	/ andel korrekte	Antal	/ andel korrekte
Opgave 1	7	100 %	7	100 %
Opgave 2	6	86 %	7	100 %
Opgave 3	7	100 %	4	57 %
Opgave 4	7	100 %	6	86 %
Opgave 5	7	100 %	7	100 %
Opgave 6	1	14 %	5	71 %
Opgave 7	7	100 %	7	100 %
Opgave 8	7	100 %	6	86 %
Opgave 9	6	86 %	7	100 %
Opgave 10	7	100 %	6	86 %
<i>I alt</i>	<i>62</i>	<i>81 %</i>	<i>62</i>	<i>81 %</i>

Tabel 9: Fejlrate for de to grænseflader fordelt på de enkelte opgaver

## A.8 Datagrundlag for subjektiv analyse

<b>Geometrisk</b>												
Spørgsmål	1	2	3	4	5	6	7	8	9	10	11	Sum
Person 1	7	8	4	8	6	2	7	3	6	6	4	61
Person 2	2	9	2	0	0	7	4	5	6	0	0	35
Person 3	6	9	9	9	9	9	9	9	9	9	9	96
Person 4	3	3	4	3	8	7	8	7	9	2	8	62
Person 5	6	7	3	2	5	7	7	6	7	8	8	66
Person 6	5	6	3	4	5	8	8	8	5	4	8	64
Person 7	4	9	4	1	5	9	7	4	9	3	9	64
M	4,71	7,29	4,14	3,86	5,43	7,00	7,14	6,00	7,29	4,57	6,57	
SD	1,80	2,21	2,27	3,44	2,88	2,38	1,57	2,16	1,70	3,26	3,36	

<b>Semantisk</b>												
Spørgsmål	1	2	3	4	5	6	7	8	9	10	11	Sum
Person 1	6	7	3	3	4	2	7	4	8	8	4	56
Person 2	6	7	3	6	5	6	3	4	6	7	6	59
Person 3	7	7	9	9	8	6	9	9	7	7	7	85
Person 4	3	3	4	5	8	7	8	8	8	7	8	69
Person 5	6	4	4	6	5	4	6	6	7	5	5	58
Person 6	5	6	3	4	6	8	6	8	6	7	8	67
Person 7	5	3	3	9	6	7	5	7	4	9	4	62
M	5,43	5,29	4,14	6,00	6,00	5,71	6,29	6,57	6,57	7,14	6,00	
SD	1,27	1,89	2,19	2,31	1,53	2,06	1,98	1,99	1,40	1,21	1,73	

Tabel 10: Datagrundlag for subjektiv analyse

## Litteratur

- [1] Perlin, K. og D. Fox. Pad: An Alternative Approach to the Computer Interface. SIGGRAPH 1993
- [2] Bederson, B.B. og J.D. Holland. Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics, ACM UIST 1994
- [3] Huynh, D.F., Drucker, S.M., Baudisch, P. og C. Wong. Time Quilt: Scaling up Zoomable Photo Browsers for Large, Unstructured Photo Collections, CHI 2005
- [4] Lance, G., Bederson, B.B, Stefik, M. og P. Baudisch. Automatic Text Reduction For Changing Size Constraints
- [5] Long, A.C., Myers, B.A, Casares, J., Stevens, S. M. og A. Corbett. Video Editing Using Lenses and Semantic Zooming,
- [6] Bier, E., Good, L., Popat, K. og A. Newberger. A Document Corpus Browser for In-Depth Reading, JCDL '04: Proceedings of the Fourth ACM/IEEE Joint Conference on Digital Libraries
- [7] Bederson, B.B., Grosjean, J., og J. Meyer. Toolkit Design for Interactive Structured Graphics, IEEE Transactions on Software Engineering, 2004
- [8] HCIL, Human-Computer Interaction Lab / University of Maryland, Piccolo Home Page, <http://www.cs.umd.edu/hcil/jazz/>
- [9] Bederson, B.B., Meyer, J., og L. Good. Jazz: An extensible Zoomable User Interface Graphics Toolkit in Java, ACM UIST 2000
- [10] Danmarks Statistik, [www.statistikbanken.dk](http://www.statistikbanken.dk), tabel NATK01, Nationalregnskab efter konto, sæsonkorrigering og prisniveau.
- [11] Danmarks Statistik, Statistisk Årbog 2004, ISBN 87-501-1397-6
- [12] Piroll, P. og R. Rao. Table Lens as a Tool for Making Sense of Data, Proceedings of the workshop on Advanced visual interface, 1993
- [13] Chin, J.P., Diehl, V.A. og K. L. Norman. Development of an Instrument Measuring User Satisfaction of the Human-Computer Interface. ACM CHI'88 Proceedings, 213-218.
- [14] Jakobsen, M. & Hornbæk, K. (2006), Evaluating a Fisheye View of Source Code, to appear in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2006).
- [15] J. Nielsen. Usability engineering, 2002, AP Professional. 23 - 48

- [16] R. Molich. User testing, Discount User Testing, Upubliceret, Copyright Dialog Design 2003, Kompendium til Datalogi V - Menneske-Datamaskine Interaktion, December 2005, Datalogisk Institut Københavns Universitet
- [17] Danmarks Statistik, [www.statistikbanken.dk](http://www.statistikbanken.dk), tabel NATK15X, ENS95, Hovedkonti pr. capita (løbende priser, kr.) efter konto og tid
- [18] Woodruff, A., Landay, J., og M. Stonebraker. Constant Information Density in Zoomable Interfaces. Proc. Advanced Vis. Interfaces '98, L'Aquila, 1998, pp. 57-65.