

On the Pagination of Complex Documents

av Anne Brüggemann-Klein, Rolf Klein, Stefan Wohlfeil

Lars-Olof Moilanen
Henrik Bäck

Översikt

- Introduktion
- Paginering
 - Dokumentmodellen
 - Sidmodellen
- Pagineringskvalité
- Pagineringsalgoritmer
- Framtiden

Introduktion

- Leseffektivitet
- Distribuering av text, figurer och tabeller
- Viktigt område
- Radbrytning
- L^AT_EX
- X_FORMATTER

Man vill uppnå leseffektivitet, slippa bläddra i onödan.

Distribuera figurer och tabeller så att de hamnar nära sina textref.

Varje paragraf måste brytas ned till rader, detta tas inte upp. Det borde funka.

Latex -> first fit, fig. tab kan hamna lite varsom

XFormatter -> denna rapports lösning på pagineringsprb.

Pagineringsmodell

Dokumentmodellen

- Informationsströmmar
 - Referenser
 - Fotnoter och figurer
- Första referens

Figurer kommer i ordning den refereras
referens i samma eller andra strömmar

Pagineringsmodellerna

Dokumentmodellen

Fotnotsström

Sed vel odio.

Textström

Lorem ipsum dolor sit amet.

Henrik
Lars-Olof

Lorem ipsum dolor sit amet,
consectetur adipiscing elit.

Maecenas iaculis, nulla ac rhoncus
luctus, lorem mauris posuere diam,
sed ullamcorper leo APPLE sit amet
nulla.

Fusce orci risus, vestibulum eu,
cursus ut, semper sit amet, enim.
Nulla id arcu sed KAU laoreet
placerat.

Figurström



Pagineringsmodell

Dokumentmodellen

Fotnotsström

Sed vel odio.

Textström

Lorem ipsum dolor sit amet.

Henrik
Lars-Olof

Lorem ipsum dolor sit amet,
consectetur adipiscing elit.

Maecenas iaculis, nulla ac rhoncus
luctus, lorem mauris posuere diam,
sed ullamcorper leo APPLE sit amet
nulla.

Fusce orci risus, vestibulum eu,
cursus ut, semper sit amet, enim.
Nulla id arcu sed KAU laoreet
placerat.

Figurström



Paginerings Sidmodellen

- En spalt
- Sidklasser
 - Enkel
 - Extra avstånd efter figurer

Höjd bredd kännetecknar en sida, alla sidor har en spalt – antas
Enkel → ett antal text eller figurströmmar i vilken ordning som helst
Extra avstånd efter figurer → Som ovan men en marginal-bottom på figurer.

Paginering

Sidmodellen



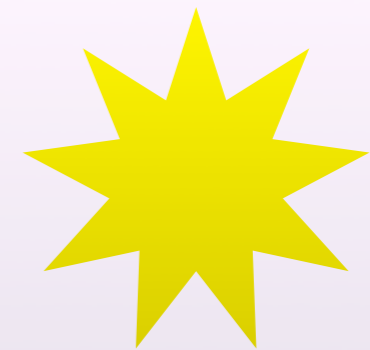
Figur 1



Figur 2

Den här sidan innehåller endast text. Hela sidans höjd kan användas av pagineringsalgoritmen.

Mellanrum beror enbart på att texterna ligger i olika stycken, texten kan fortsätta



Figur 3

på en annan sida, där det exempelvis kan finnas en figur.

Pagineringskvalité

- **Plass:**

$$Lin(P) = \sum_{i=1}^n \sum_{j=1}^{r(i)} |P(f_i) - P(R_j(f_i))|$$

$$Quad(P) = \sum_{i=1}^n \sum_{j=1}^{r(i)} (P(f_i) - P(R_j(f_i)))^2$$

- **Brüggemann-Klein, Klein, Wohlfeil:**

$$Turn.S(\alpha, \beta, P) = \beta(p-1) + \sum_{i=1}^n \alpha(P(f_i) - P(R(f_i)))$$

$$Turn.D(\alpha, \beta, P) = \beta(S(p)-1) + \sum_{i=1}^n \alpha(S(P(f_i)) - S(P(R(f_i))))$$

Badnessfunktioner

Plass: NP-komlett

$r(i)$ är antalet referenser till f_i

$R_j(f_i)$ är boxen med j -te referensen till f_i

Hög kvalité = lågt fn-värde

α, β summas till 1

Pagineringsalgoritmer

- Mål
 - Behålla ordning
 - Inga horungar
 - Fylla sidor
 - Placering av objekt

textordning
enkel och dubbel horunge
100% (90%)

Pagineringsalgoritmer

On-line paginering

- Ingen kunskap om hela dokumentet
- Sämre resultat
- “Figure First Fit”

Pagineringsalgoritmer

Exempel - On-line paginering

figur	f_1	f_2	f_3	...	f_n	f_{n+1}	f_{n+2}
höjd	1	1	3		3	1	1

Sidhöjd: 5

Tomrum efter figur: 1

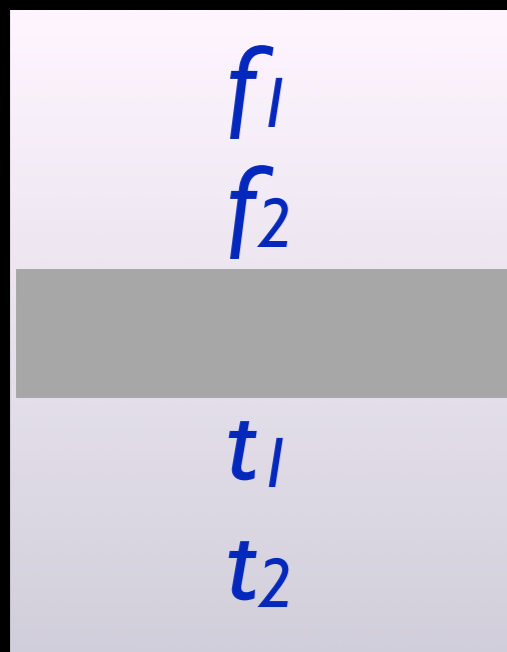
$$R(f_1) = t_1$$

$$R(f_2) = t_2$$

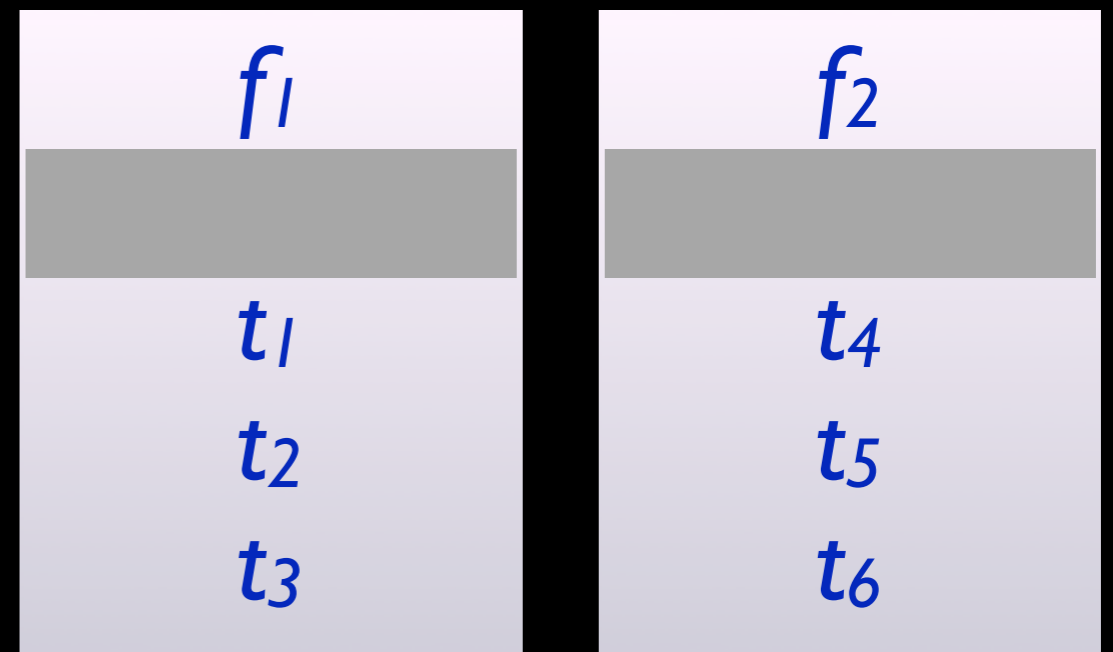
Pagineringssalgoritmer

Exempel - On-line paginering

Alternativ 1



Alternativ 2

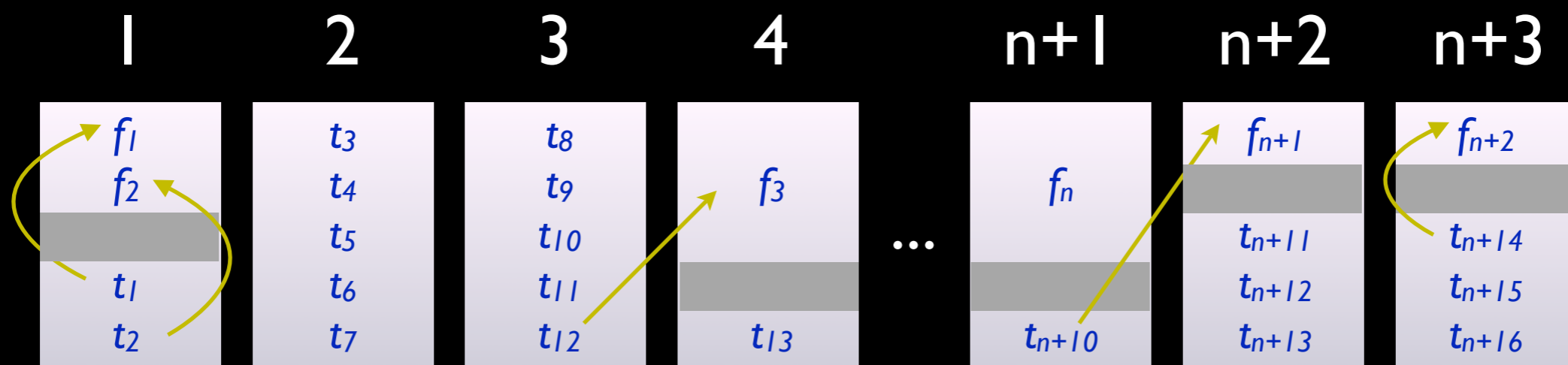


Fler än 2 alt. men sämre utgång

Pagineringsalgoritmer

Exempel - On-line paginering

Alternativ I



$$R(f_i) = t_{i+9} \mid \forall i \in [3, n+1]$$

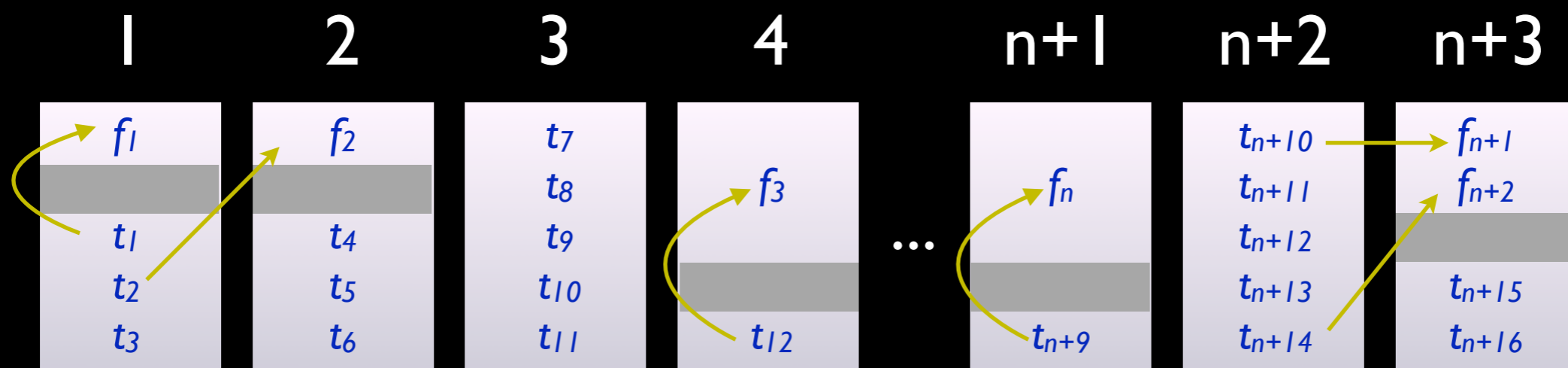
$$R(f_{n+2}) = t_{n+14}$$

f₁, f₂ och f_{n+2} på "rätt" sida, de andra n-1 figurerna en sida efter refererande textblock.
Badness n-1

Pagineringsalgoritmer

Exempel - On-line paginering

Optimal I

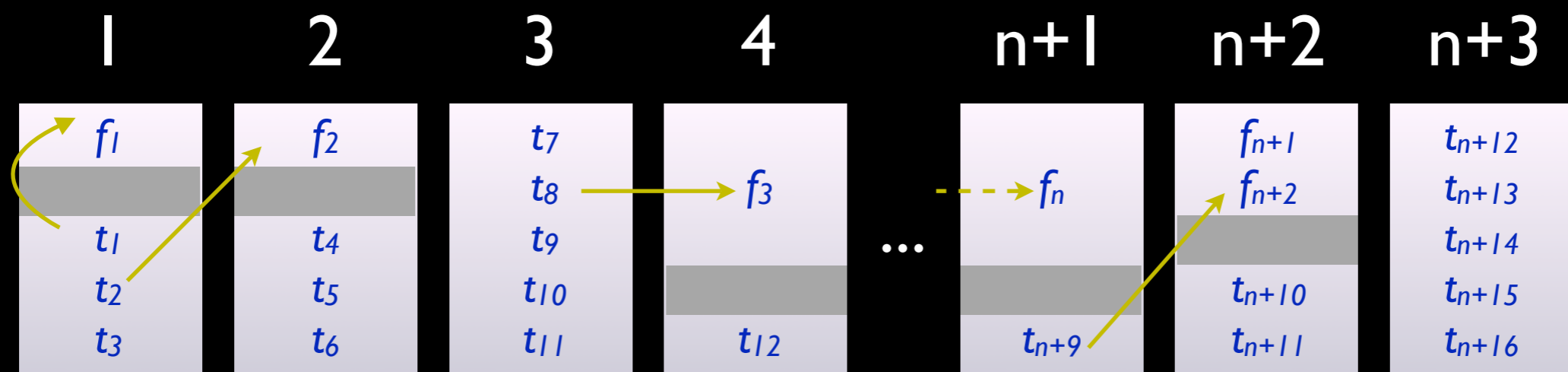


Badness 3 (?)

Pagineringsalgoritmer

Exempel - On-line paginering

Alternativ 2

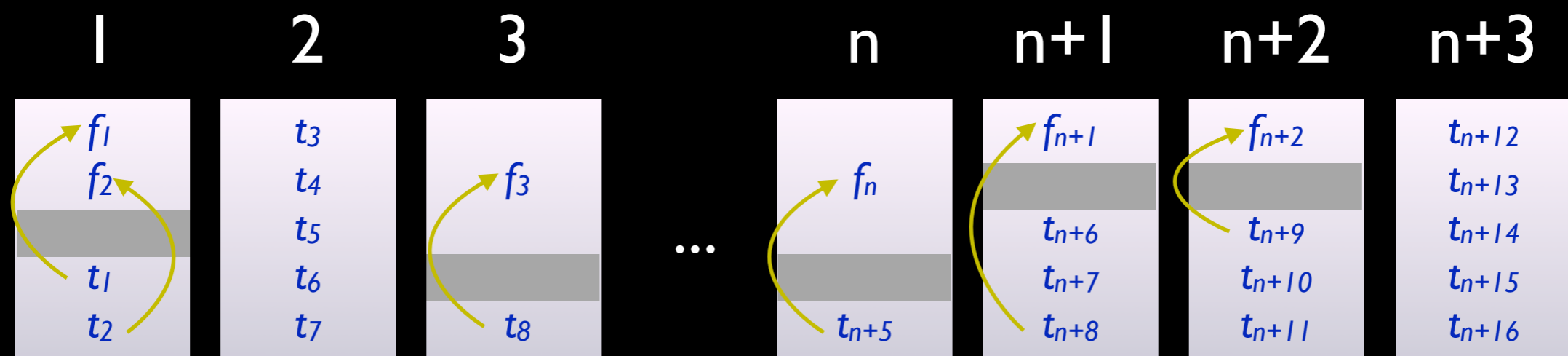


$$R(f_i) = \begin{cases} t_{i+5} & | \quad \forall i \in [3, n] \\ t_{i+7} & | \quad i \in \{n+1, n+2\} \end{cases}$$

Pagineringsalgoritmer

Exempel - On-line paginering

Optimal 2

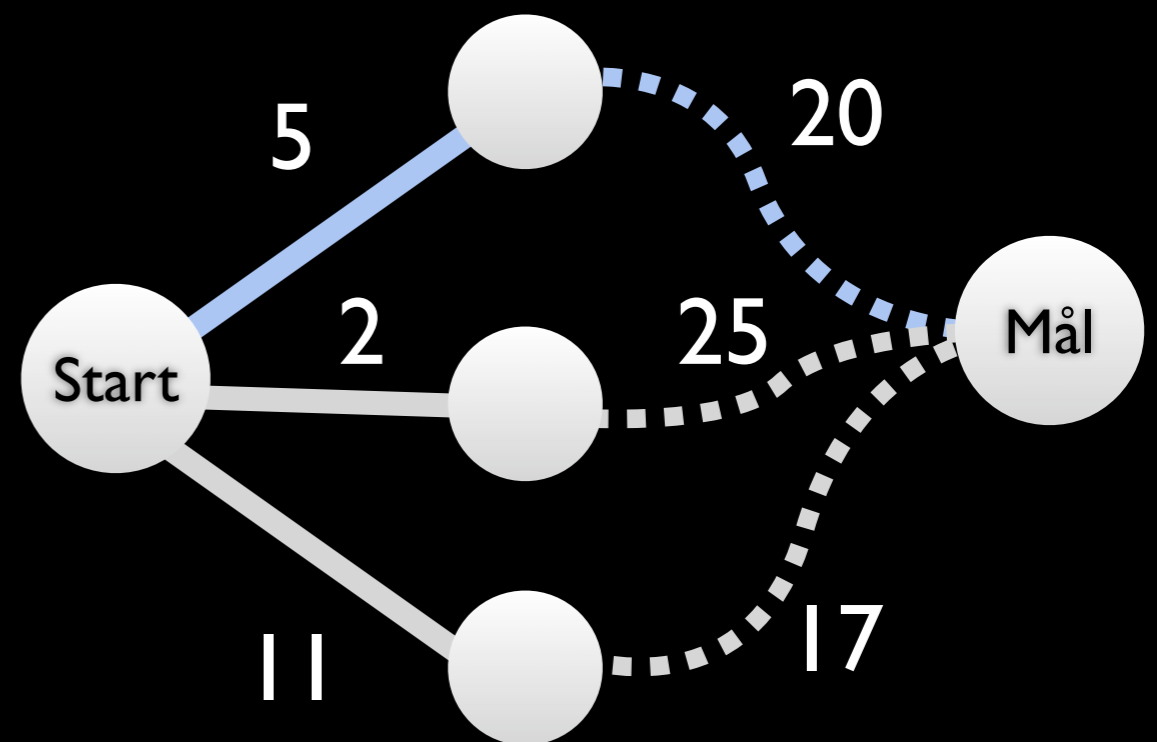


Badness: 0 (?)

Pagineringsalgoritmer

Dynamisk programmering

- Bevis för att *Turn.S* och *Turn.D* tillåter effektiv optimering
- Algoritm
- Optimeringsprincipen



Exempel från Wikipedia

optimala lösningen = optimala lösningen av subproblem
Richard Bellman, 1940-talet
ordet programmering har en matematisk innebörd (optimering)
sparar lösningen

Pagineringsalgoritmer

Dynamisk programmering

- Enkelsidigt
 - $Turn.S(\alpha, \beta, P_{k,l})$ partiell paginering för k textrutor och l figurrutor.
 - Hängande referens till figur $f_w \Rightarrow$
 $Turn.S(\alpha, \beta, P_{k,l}) += p+1 - P_{k,l}(R(f_w))$
 - Om $Turn.S(\alpha, \beta, P_{k,l})$ är optimal, är $Turn.S(\alpha, \beta, P_{i,j})$ där $P_{i,j}$ motsvarar radering av sista sidan i $P_{k,l}$ också optimal.

Pagineringsalgoritmer

Dynamisk programmering

- Dubbelsidigt
 - $Turn.D(\alpha, \beta, P_{k,l})$ partiell paginering för k textrutor och l figurrutor.
 - Hängande referens till figur $f_w \Rightarrow$
 $Turn.D(\alpha, \beta, P_{k,l}) += p+1 - P_{k,l}(R(f_w))$ om p inte är en vänstersida

Pratar inte längre om DEN optimala lösningen $P_{i,j}$ då det kan finnas två.

Pagineringsalgoritmer

Praktiska resultat

Paginering	<i>Turn.S</i>				<i>Turn.D</i>			
	doc1	doc2	doc3	doc4	doc1	doc2	doc3	doc4
Författarens	30	72	-	72	13	37	-	39
“First-fit”	39	70	55	56	17	38	25	26
<i>Turn.S</i> optimal	31	65	47	52				
<i>Turn.S</i> 90% optimal	-	59	38	47				
<i>Turn.D</i> optimal					13	30	25	25
<i>Turn.D</i> 90% optimal					10	26	18	23

doc1 – rapport för regeringen i Bayern

doc2 – bokkapitel från Software Engineering

doc3 – doktorsavhandling ang. skolutvecklingsforskning i Tyskland

doc4 – kapitel från en bok ang. algoritmer för att lösa gemoetriproblem

Framtiden

- Nuvarande system inte optimala
- Metoden har stor potential
- Inkludering av algoritmer i existerande typsättningsystem?
- Tillfredsställa typsättare istället för datavetare.