

OCR error correction of Swedish printed texts at Språkbanken Text

Dana Dannélls
Språkbanken Text
Institutionen för svenska, flerspråkighet
och språkteknologi

OCR fel

Antal träffar: 293

« < 1 2 3 4 5 6 7 8 9 10 11 12 > »

Gå till sida av 12 [Visa kontext](#)

JUBILEUMSARKIVET (PILOT)

Voisins Sedan, elegant ooh smakfull.

här hemma oftast ej särdeles » myckel om vårt lands ställning nr rent finternationell synpunkt — rå horisont år trång ooh vår plats undanskymd.

Vårt turistväsen har ingalunda hört till världens bäst organiserade, ooh det är därför också med den största till

Danna fråga är av största betydelse ooh räckvidd ieko endast ur ren turist- «ynpunkt, ty dan sammanhänger även nod hela vårt näringsliv.

gäller den fond av vetands och kunskap om Sverige, som än mer skall öppna världens blickar för våra naturrikedomar ooh kvalitets * varor, det gäller kort - — ooh fult sagt — profiten.

om Sverige, som än mer skall öppna världens blickar för våra naturrikedomar ooh kvalitets * varor, det gäller kort - — ooh fult sagt — profiten.

ligen • — » n stunds tågväntan i en si.-t- Itionshall torde övertyga —, att här [kunnat göras starka ting, bilder som Isega ooh fånga, beror helt på den gra- Viska konstens höga nivå i Sverige.

> — > främlingen till den stora utställningen i Göteborg, man ser och beundrar vad Sverige hjr att bjuda på, vad det är ooh vad det förmår.

ionen grupperar sig kring en väldig horisontell, sju meter lång turistkarta av vårt avlänga land, varpå dess sevårdheter ooh kommunikationer gjorts så åskådliga som möjligt .. Ds större turistorterna äro utmärkta med miniatyrmodeller av hus ooh anläggning

kommunikationer gjorts så åskådliga som möjligt .. Ds större turistorterna äro utmärkta med miniatyrmodeller av hus ooh anläggningar, på järnvägslinjerna gå miniatyrtåg och på vattnen båtar.

t Man vandrar härefter gärna rundi | utefter utstÄhningshallens av fotogra-, fier fyllda väggfält, ooh man finner i snart, att materialets vidd mor än dosa; enskildheter innebär dess instruktiva | betydelse, penna med stor omsorg ut!

Landskap för landskap presenterar sig, vart ooh ett vill, visa sin karaktär ooh sitt skaplynm | En svortgekäni are vet, att de alla äro.

Landskap för landskap presenterar sig, vart ooh ett vill, visa sin karaktär ooh sitt skaplynm | En svortgekäni are vet, att de alla äro.

nå- I got annat likt — ty den är svensk j Svensk naturs outtömliga skönhetsrike * I dom låter en var tjusas av sitt ideal | ooh längta bort till de traktor, dit hå- I gen står.

Historiska ooh konstnärliga j minnen, nya tiders frammarsch med I storartade anläggningar, storstad ooh I småstadsidyU uppfriska ooh tillfreds » I stÄ

Historiska ooh konstnärliga j minnen, nya tiders frammarsch med I storartade anläggningar, storstad ooh I småstadsidyU uppfriska ooh tillfreds » I stÄlla här varje interesse.

n konstnärliga j minnen, nya tiders frammarsch med I storartade anläggningar, storstad ooh I småstadsidyU uppfriska ooh tillfreds » I stÄlla här varje interesse.

I Stockholm och Göteborg ha fått va: I sitt väggfält, ooh att Stockholm i stolt I suveränitet där tar priset är ju också | ganska naturligt Att våra fjälltrak * I ter ägnats betydande utrymmen härj

ooh det hf svårt att tänka i ' sig huru man på utatlihi; n; - on i tanken kunde ha förts landet närmaro.

* • alla delar skola öppnas, otalig » nya trä- i dar spinnaa, nya intressen väckas ooh omsättas i betydelsefull handling.

lott gå till en på ett bord intill placerad handskreven liggare, avskrift av justitiedepartementets omröstningspro- tokoll ooh upptagande siffrorna för varje omröstningsdistrikts antal röstberättigade, röstande och giltiga ja- och nejsedlar samt ogilla som avgivi

Dit gå förvisso tankarna under dagens tunga och hetta, ooh dit föra stegen så lätt på kvällskröken.

De äro däremot rätt talrikt företrädda efter midnatt på Avenyn och Östra hamm gatan, ooh de razzior som redan företagits behöva nog upprepas.

ed direktör Gustaf Bergman på Stora teatern att han skall övertaga ledningen av teatern samt spela för nu till att qvist ooh Daniel Berg, vilka tidigare utmärkt sig på revyernas område.

publicerat ett stort antal artiklar om ' Sverige ooh svenska förhållanden, ägna- ' de att reda upp utlänningarnas begrepp | om Sverige och.

OCR fel

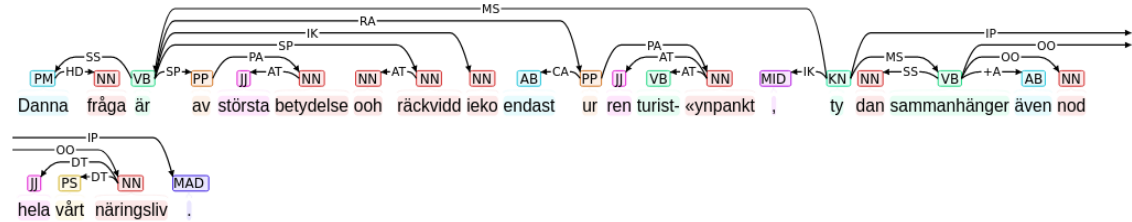
Antal träffar: 293

« < 1 2 3 4 5 6 7 8 9 10 11 12 > » Gå till sida av 12 Visa kontext

här hemma oftast ej särdeles » mycket om vårt lan

gäller den fond av vetands och kunskap om Sverige
om Sverige, som än mer skall öppna världens blick
ligen • — » n stunds tågväntan i en si.-t- Itionshall to
> — > främlingen till den stora utställningen i Göte
ionen grupperar sig kring en väldig horisontell, sju
ommunikationer gjorts så åskådliga som möjligt ..
t Man vandrar härefter gärn

Dependensträd



Landskap för landskap presenterar sig, vart ooh ett vill, visa sin karaktär ooh sitt skaplynm | En svortgekäni are vet, att de alla äro.

Landskap för landskap presenterar sig, vart ooh ett vill, visa sin karaktär ooh sitt skaplynm | En svortgekäni are vet, att de alla äro.

nå- I got annat likt — ty den är svensk | Svensk naturs outtömliga skönhetsrike * I dom låter en var tjasas av sitt ideal | ooh längta bort till de traktor, dit hå- I gen står.

Historiska ooh konstnärliga j minnen, nya tiders frammarsch med I stortade anläggningar, storstad ooh

Historiska ooh konstnärliga j minnen, nya tiders frammarsch med I stortade anläggningar, storstad ooh | småstadsidyU uppfriska ooh tillfreds » I ställa här varje intresse.

1 konstnärliga j minnen, nya tiders frammarsch med I stortade anläggningar, storstad ooh | småstadsidyU uppfriska ooh tillfreds » I ställa här varje intresse.

I Stockholm och Göteborg ha fått va: I sitt väggfält, ooh att Stockholm i stort I suveränitet där tar priset är ju också I ganska naturligt Att våra fjällt

ooh det hf svårt att tänka i ' sig huru man på utatlihi; n; - on i tanken kunde ha förts landet när

* • alla delar skola öppnas, otalig » nya trä- i dar spinnaa, nya intressen väckas ooh omsättas i betydelsefull handling.

lott gå till en på ett bord intill placerad handskreven liggare, avskrift av justitiedepartementets omröstningspro- tokoll ooh upptagande siffrorna för varje omröstningsdistrikts antal röstberättigade, röstande och gi

Dit gå förvisso tankarna under dagens tunga och hetta, ooh dit föra stegen så lätt på kvällskröken.

Ett första steg mot en fri resurs för OCR

- Ansökan med språkbanken Text (SBX) och universitetsbibliotek (UB)
 - UB: Eric Siira, Lennart Stark
 - SBX: Lars Borin, Gerlof Bouma, Dana Dannélls
- Syftet var att ta ett första steg mot att skapa en nationell resurs för OCR-tolkning av tryckta texter med fokus på fraktur
- Resursen ska kunna utnyttjas av bibliotek, museer och arkiv
- Financieras av Vinnova
 - september 2013 -- augusti 2014

SPRÅKBANKENTEXT

VINNOVA
Sweden's Innovation Agency

En fri molntjänst för OCR: Prototypbyggnad

- Material
 - Urval av vetenskapliga utgåvor av äldre texter samt äldre frakturtyck
 - Ordlistor som täcker ordförådet under 1800-talet

Digitaliserade versioner av vetenskapliga utgåvor av äldre texter

Litteraturbanken
Hesselman/Lamm
1910–12

THEN SWÄNSKA

A R G U S

N:o I.

Omne tunc punctum qui miscuit
utile dulci.

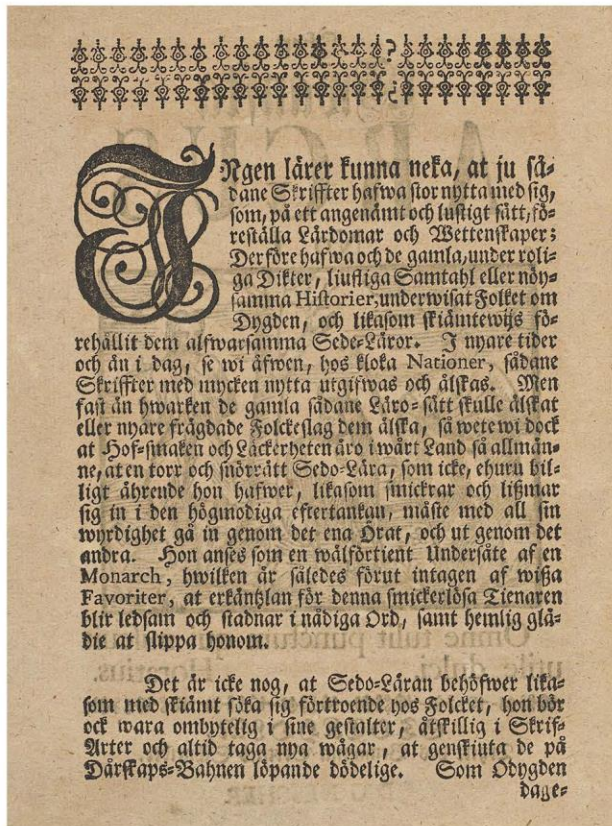
Horatius.

Ingen lærer kunna neka, at ju sådane Skrifter hafwa stor nytta med sig, som, på ett angenämt och lustigt sätt, föreställa Lärdomar och Wettenskaper; Derföre hafwa och de gamla, under roliga Dikter, liufliga Samtahl eller nöysamma Historier, underwisat Folket om Dygden, och likasom skiämtewijs föhållit dem alfwarsamma Sede-Läror. I nyare tider, och än i dag, se wi äfwen, hos kloka Nationer, sådane Skrifter med mycken nytta utgifwas och älskas. Men fast än hwarken de gamla sådane Läro-sätt skulle älskat eller nyare frägdade Folckeslag dem älska, så wete wi dock at Hof-smaken och 10 Läckerheten äro i vårt Land så allmänne, at en torr ock snörrätt Sedo-Lära, som icke, ehuru billigt ährende hon hafwer, likasom smickrar och lissmar sig in i den högmodiga efter-tankan, måste med all sin wyrdighet gå in genom det ena Örat, och ut genom det andra. Hon anses som en wälförtient 16 Undersåte af en Monarch, hwilken är således förut intagen af wissa Favoriter, at erkantzlan för denna smickerlösa Tienaren blir ledsam och stadnar i nådiga Ord, samt hemlig glädie at slippa honom.

Det är icke nog, at Sedo-Läran behöfwer likasom med 20 skiämt söka sig förtroende hos Folcket, hon bör ock wara

Titel Hs Frispråkaren. *af str:* Sambling af Frja Tankar från Utopia.
7 *Uz* kloka Folckeslag 8-10 *Uz* älskas: Wi wete, at Hofsmaken 11 *Hs* är
11 *Hs* allmän 12-13 *Uz* icke lika som smickrar 14 *Uz* med alt sit wärde
1.

Digitaliserade versioner av äldre frakturtryck



S ingen lærer kunna neka, at ju så-
dane Skrifter hafwa stor nytta med sig,
som, på ett angenämt och lustigt sätt, för-
rethålla Lärdomar och Wettenskaper;
Der före hafwa och de gamla, under costi-
ga Dikter, liusliga Samtahl eller nöys-
samma Historier, underwisat Folket om
Dygden, och liksom skäntewijs för-
rethållit dem allswariga Sedo-Lärer. I nyare tider
och än i dag, se wi äfwen, hos floka Nationer, sådane
Skrifter med mycken nytta utgifwas och ästas. Men
fast än hwarken de gamla sådane Läro-sätt skulle ästas
eller nyare frågdade Folckeslag dem ästka, så wete wi dock
at Hof-smaken och Lækerbeten äro i vårt Land så allmän-
ne, at en torr och sinderätt Sedo-Lära, som icke, ehuru bil-
ligt ährende hon hafwer, liksom smickrar och lifnjar
sig in i den högmotiga eftertankan, måste med all sin
würdighet gå in genom det ena Drat, och ut genom det
andra. Hon anses som en wälsförrent Underfåte af en
Monarch, hwilken är således förut intagen af wisja
Favoriter, at erkänslan för denna smickerlösa Dienaren
blir lebsam och stadnar i nädiga Ord, samt hemlig glä-
die at slippa honou.

Det är icke nog, at Sedo-Läran behöfwer lika-
som med skänt sika sig förtroende hos Folcket, hon bör
ock wara ombytellig i sine gestalter, ästfillig i Skrif-
Arter och altid tagga nya wägar, at genskiuta de på
Därkaps-Bahnen löpande dödelige. Som Ddygden
dager

Then Swänska Argus
1732

Data

Typ	Namn	Tidsperiod	Tecken	Ord	Beskrivning
Text	Svensk fraktur	1626–1816	282,432	48,112	199 pages
Text	Then swänska Argus	1732–1734	259,468	43,640	177 pages
Lexicon	Swedberg	1700		16,158	utan böjningar
Lexicon	Dalin	1800		509,924	Med böjningar

OCR-system

Öppen källkod OCR-motorer

- Tesseract med blandade typsnitt, färdiga modeller för olika språk, stöd genom ordlistor, mindre bra fraktur model för svenska
- OCRopus, höggradigt konfigurerbar arkitektur, ingen språkmodel för svenska

Kommersiella OCR-motorer

- Abbyy FineReader, färdiga modeller för olika språk, stöd genom ordlistor, mindre bra fraktur model för svenska

Träna teckenmodeller med OCRopus

ingen inskränktning, der alla äro dygdiga. Men wi böjas ofta til



§. 2. En förmän, som af menniskjor så högt älskas, behöfver



I sjukdomarnas annaler synas kolerans såväl sym-

I sjukdomarnas annaler synas kolerans såväl sym=

tomer som härjningar lämpligast kunna jämföras med di-

tomer som härjningar lämpligast kunna jämföras med di=

gerdödens. ? Båda farfoterna bära nästan samma färger.

gerdödens. Båda farfoterna bära nästan samma färger.



Analys på teckennivå

derfätare i det andras Länder, i anseende til lösa egendomen, i så
måtto: At om någon Franssk undersäte genom döden afgår i Swe-
rige eller tilhörande Provincer, så skal hans rätta arfwingar, Te-
stamentarier eller Legatarier, eller ho det wara må, som hafwa rät-
tighet til arfwet, antingen han bor i Riket eller är främmande wa-
ra obetaqit at niuta all den dödas här lemnade egendom

Tesseract

derföre tro i det undrad Fänder, i en liten de fri lof, i tusendom til, i fä-
dersa täre i det andras Länder, i anseende til lösa egendomen, i fil

llen klokt Folk om någon Njt Bull' under fält Bendin boden utgår i Cede-
nätte: Ate om någon Fraeffe under Esäte genom döden afgår i Swe-

tigt eller tilhörande Öron intet, få fögi hund kätter ur flollgur, Tr.
rige eller tilhörande Provincer, så ffal hans rätta arfwingar, Te-

Hennes be Eskier eller Fru Bielkior, eller ho bel trutet mer, som härmed tal-
siamel tc ricr eller Legatarier, eller ho det wara ma, som hafwa rät-

ett till arfwet, antingen han bor i Riket eller är främmande wa-
liqrl till arfwet, antingen han bor i Riket eller är främmande wa-

10 obet #811 ill lillelse: Bll Den dödelig häll löning be egendom, til nu af-
ra obet agit ate njuta all den dödas här lemnade egendom, uta af-

OCRopus

derföre tro i det undrad Fänder, i en liten de fri lof, i tusendom til, i fä-
dersa täre i det andras Länder, i anseende til lösa egendomen, i fil

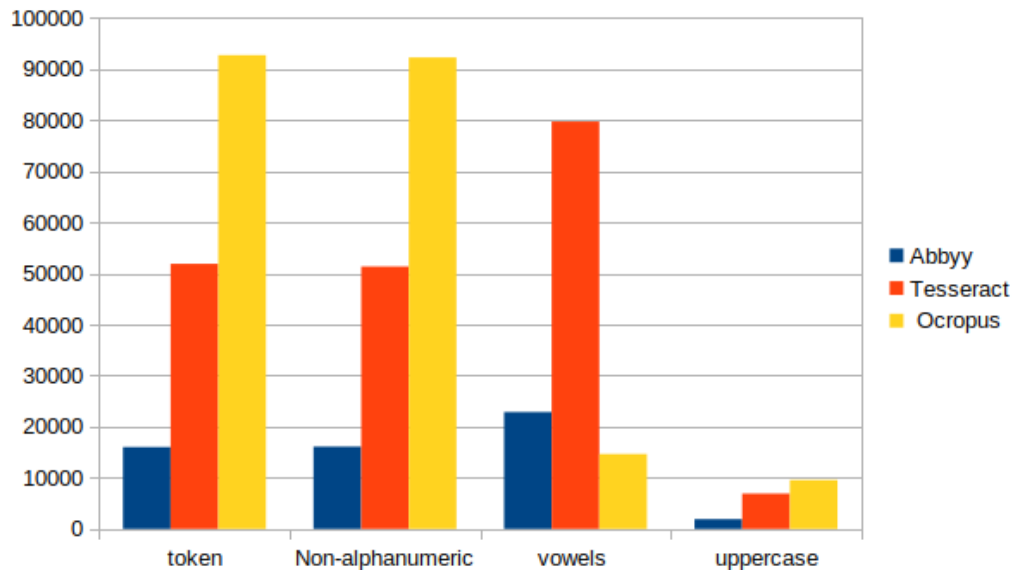
uatto: At om någon Franfk underfäle genom döden afgår i Swe- ce: 18,
mätto: At om någon Franff undersäte genom döden afgår i Swe-

rige eller tilhörande Provincer, så skal hans rätta arfwingar, Te-
rige eller tilhörande Provincer, så ffal hans rätta arfwingar, Te-

ftanmo tä rier eller Legatarior, eller ho det wara må, som hafwa rät-
siamel tc ricr eller Legatarier, eller ho det wara ma, som hafwa rät-

tighel til arfwet, antingen han bor i Riket eller är främmande wa-
liqrl till arfwet, antingen han bor i Riket eller är främmande wa-

OCR error analys



Feature	Abbyy	Tesseract	Ocropus
token count	15992	51852	92690
non-alphanumeric characters	16102	51350	92231
vowels	22871	79714	14645
uppercase letters	1904	6891	9556

Table 1. Analysis of the tokens that were recognized by each OCR system.

Evalueringresultat

Data	Engine	CER (%)	WER (%)
Argus	Off-the-shelf OCRopus	10.4	47.2
	Off-the-shelf Tesseract	39.1	76.9
	Modified Tesseract	48.1	NA
	Modified OCRopus	11.7	49.1
Mixed samples	Off-the-shelf OCRopus	14.0	53.8
	Off-the-shelf Tesseract	37.6	72.3
	Modified Tesseract	40.2	NA
	Modified OCRopus	16.8	59.7

OCR post-processing med ordlistor



allamedhstärsta

-ol: "allamedzstärsta "

+ol: "alla medh största"

ickeallenastför

-ol: "icke allenastför"

+ol: "icke allenast för"

Thekommoochtill

-ol: "thekommoochtill"

+ol: "Them kommo och till"



nampnkuunnige => Nu Kunnige

instead of: nampn kunnige

KarinsSundh? => Darius Stundh ?

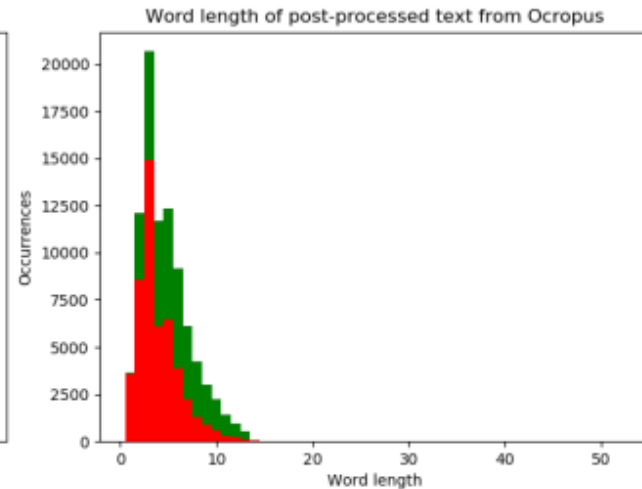
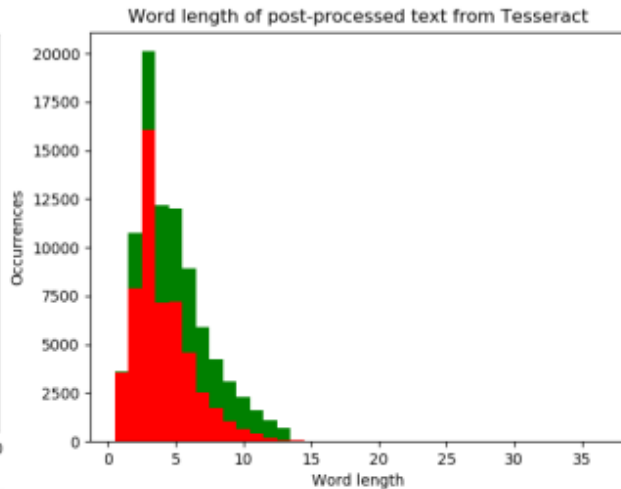
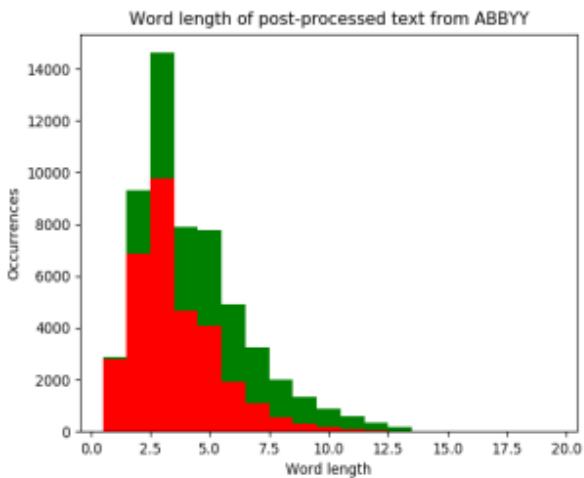
instead of: Karins Sundh

lingtrppythiRöda => långt Kropp vthi Röda

instead of: långt upp uthi röda

Hur mycket täcker våra lexikon?

.



Molntjänsten

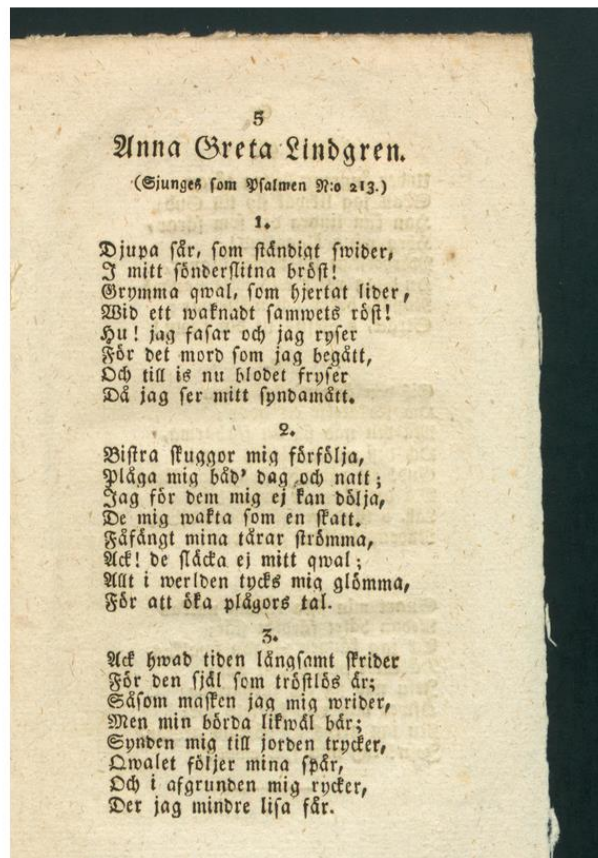
Ordlister leder inte alltid till förbättring.

Svårt att hitta heltäckande lexikon för olika språk och tidsperioder.

Modellen är tränad på relativt liten data med manuell transkription.

Med post-processing tar det väldigt lång tid.

<https://spraakbanken.gu.se/projekt/ocr-cloud>



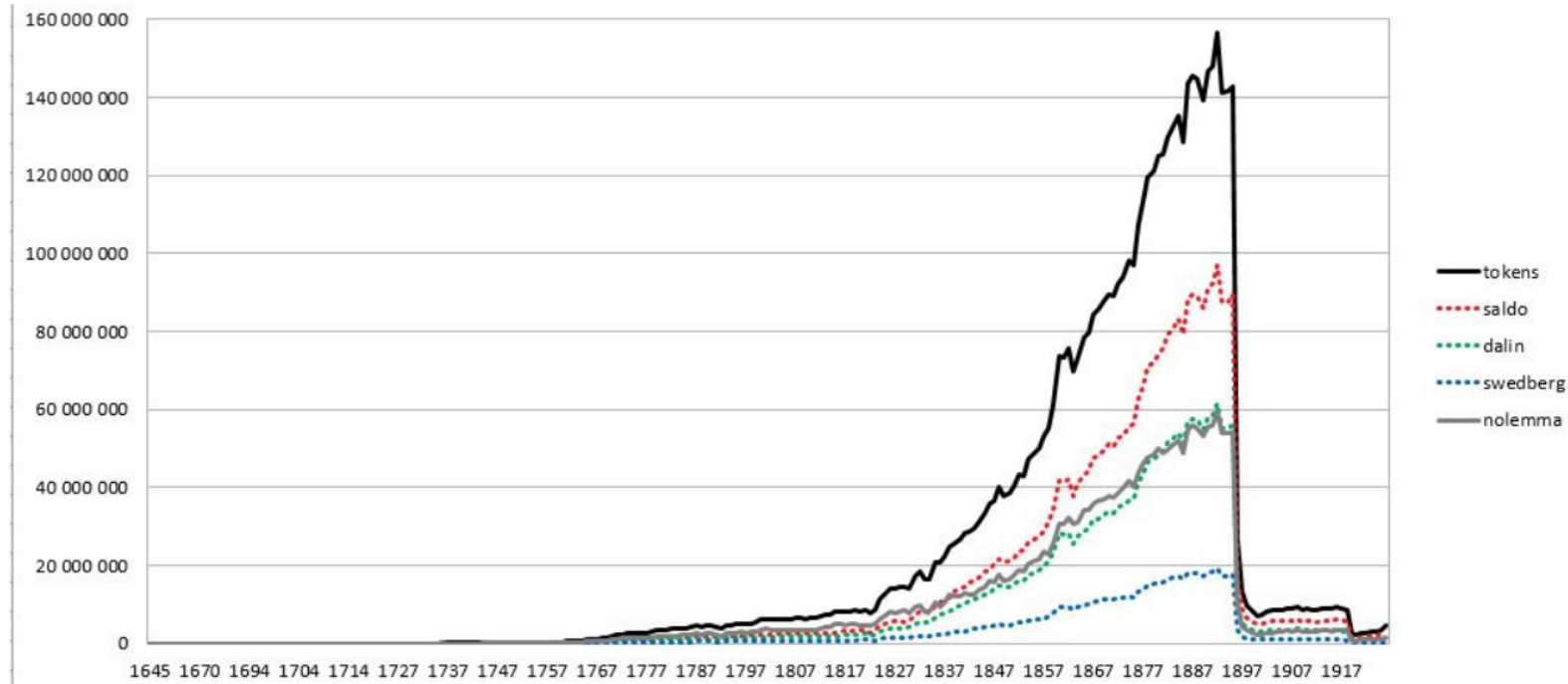
Download Text

Download HTML

Anna Greta Lindgrn.
(Sjunges som Lfalmen Mw 13.)
Djupa får, som ftåndiat fwider,
I mitt fõnderlitna brõftl
Erymma gwal, som hjertat lider,
Aid ett waknadt famwets röftl
åHu l jag fajar och jag ryfer
Rõr det mord som jag begått,
Odoch till is nu blodet fryfer
lå jag fer mitt fyndamått.

~.
Biftra fkuggor mig förfõlja,
Hlåga mia bäck dag ,och natt;
ag för dem mig ej kan dõlja,
lDe mig wakta som en fkatt.
Råfångt mina tårar ftrõmma,
2Ack; de flåcka ej mitt gwal r
Eittl i werlden tycks mia glõmma-
För att öka plågors tal.
~lek hwad tiden lånafamt fkri-
der
IFör den fjål som trõftlös år;
Såfom malken jaa mig wrider,
lMen min börda likwål bår;
Synden mig till jorden trycker,
l walet följer mina spår,
Och i afgrunden mig rycker,
lDer jag mindre lifa får.

Hur mycket täcker våra lexikon? -- Kubhist 2



Digitisering av tidningsmaterial



```
<SP ID="SP126" WIDTH="0" HPOS="3546" VPOS="4781"/>
<String HPOS="3571" VPOS="4777" WIDTH="72" HEIGHT="23" CONTENT="Dina" STYLEREF="style8" WC="0.36"/>
<SP ID="SP127" WIDTH="0" HPOS="3643" VPOS="4783"/>
</TextLine>
- <TextLine ID="Line25" HPOS="2962" VPOS="4066" WIDTH="683" HEIGHT="770">
  <String HPOS="3065" VPOS="4807" WIDTH="98" HEIGHT="27" CONTENT="läppar" STYLEREF="style8" WC="0.38"/>
  <SP ID="SP128" WIDTH="0" HPOS="3163" VPOS="4815"/>
  <String HPOS="3188" VPOS="4809" WIDTH="47" HEIGHT="23" CONTENT="äro" STYLEREF="style8" WC="0.62"/>
  <SP ID="SP129" WIDTH="0" HPOS="3235" VPOS="4815"/>
  <String HPOS="3262" VPOS="4814" WIDTH="58" HEIGHT="18" CONTENT="som" STYLEREF="style8" WC="0.71"/>
  <SP ID="SP130" WIDTH="0" HPOS="3320" VPOS="4815"/>
  <String HPOS="3345" VPOS="4809" WIDTH="101" HEIGHT="26" CONTENT="skogar" STYLEREF="style8" WC="0.39"/>
  <SP ID="SP131" WIDTH="0" HPOS="3446" VPOS="4816"/>
  <String HPOS="3471" VPOS="4809" WIDTH="71" HEIGHT="25" CONTENT="efter" STYLEREF="style8" WC="0.43"/>
  <SP ID="SP132" WIDTH="0" HPOS="3542" VPOS="4817"/>
  <String HPOS="3567" VPOS="4815" WIDTH="73" HEIGHT="21" CONTENT="regn" STYLEREF="style8" WC="0.51"/>
  <SP ID="SP133" WIDTH="0" HPOS="3640" VPOS="4826"/>
</TextLine>
- <TextLine ID="Line26" HPOS="2962" VPOS="4066" WIDTH="683" HEIGHT="798">
  <String HPOS="3065" VPOS="4839" WIDTH="22" HEIGHT="21" CONTENT="2" STYLEREF="style8" WC="1.00"/>
  <SP ID="SP134" WIDTH="0" HPOS="3087" VPOS="4856"/>
  <String HPOS="3129" VPOS="4839" WIDTH="39" HEIGHT="24" CONTENT="Ur" STYLEREF="style8" WC="0.66"/>
  <SP ID="SP135" WIDTH="0" HPOS="3168" VPOS="4846"/>
  <String HPOS="3208" VPOS="4839" WIDTH="143" HEIGHT="23" CONTENT="" STYLEREF="style8" WC="0.67"/>
  <SP ID="SP136" WIDTH="0" HPOS="3351" VPOS="4839"/>
  <String HPOS="3392" VPOS="4840" WIDTH="120" HEIGHT="23" CONTENT="(William" STYLEREF="style8" WC="0.68"/>
  <SP ID="SP137" WIDTH="0" HPOS="3512" VPOS="4847"/>
  <String HPOS="3533" VPOS="4840" WIDTH="91" HEIGHT="23" CONTENT="Shake" STYLEREF="style8" WC="0.60" SUBS_
SUBS_TYPE="HypPart1"/>
<HYP CONTENT="-"/>
</TextLine>
- <TextLine ID="Line27" HPOS="2962" VPOS="4066" WIDTH="683" HEIGHT="833">
  <String HPOS="3066" VPOS="4871" WIDTH="110" HEIGHT="27" CONTENT="speare" STYLEREF="style8" WC="0.49" SUBS
```


Utvärdering och anpassning av en förbättrad OCR-process vid massdigitalisering

- Ett samarbete med SBX och KB
 - SBX: Dana Dannélls
 - KB: Lars Björk, Torsten Johansson
 - Zissor: Ove Dirdal
- Syftet är att genomföra en formell utvärdering av en två-model metod med målet att det ska implementeras i KB:s massdigitaliseringsprocess för dagstidningar
- Financieras av RJ
 - januari 2019 -- augusti 2021



SPRÅKBANKEN **TEXT**

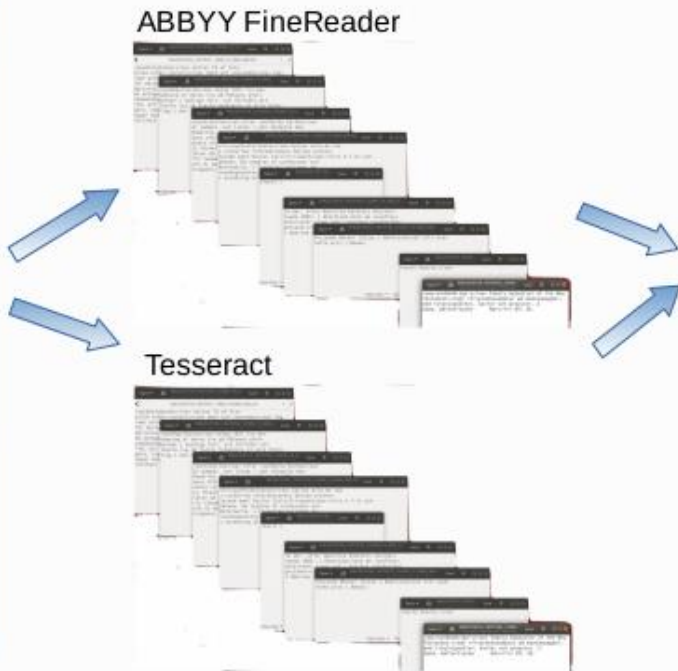


SWE-CLARIN



National Library
of Sweden

Två-model metoden



```
<div data-bbox="195 295 285 445" data-label="Text" style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Kangorese  
S  
...  
</div>  
<div data-bbox="235 455 285 505" data-label="Text" style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Kangorese  
...  
</div>  
<div data-bbox="235 515 285 675" data-label="Text" style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">S  
...  
</div>
```


Trafik västerut sedan Norge är fritt.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Trafik västerut sedan Norge är fritt. Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Trafik västerut sedan Norge är fritt. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Dansk dog av glädje.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Dansk dog av glädje. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Hilberdocka fick inte utställas.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Hilberdocka fick inte utställas. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Tysk rovdrift Norges skogar.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Tysk rovdrift Norges skogar. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Skarp kritik mot jordkontroll.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Skarp kritik mot jordkontroll. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

BILRUTOR

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. BILRUTOR. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Åtvidabergs hängmappar och mappskåp

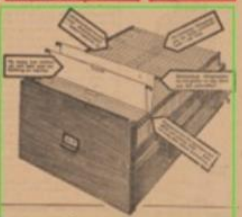
Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Åtvidabergs hängmappar och mappskåp. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.



AFTONBLADET
Årens 1932
MORNINGTIDNING
HELVÄMMER
Åtvidabergs

Leopold inom försäkrade fönster.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Leopold inom försäkrade fönster. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.



Åtvidabergs
Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Trafik västerut sedan Norge är fritt.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Trafik västerut sedan Norge är fritt. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Dansk dog av glädje.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Dansk dog av glädje. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Hilberdocka fick inte utställas.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Hilberdocka fick inte utställas. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Tysk rovdrift Norges skogar.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Tysk rovdrift Norges skogar. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Skarp kritik mot jordkontroll.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Skarp kritik mot jordkontroll. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

BILRUTOR

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. BILRUTOR. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

Åtvidabergs hängmappar och mappskåp

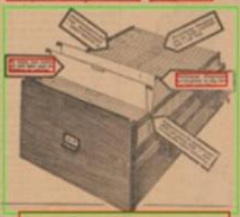
Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Åtvidabergs hängmappar och mappskåp. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.



AFTONBLADET
Årens 1932
MORNINGTIDNING
HELVÄMMER
Åtvidabergs

Leopold inom försäkrade fönster.

Störst i västerut på "gula rymmer" — Årens inkomst i Norge. Leopold inom försäkrade fönster. Störst i västerut på "gula rymmer" — Årens inkomst i Norge.



Åtvidabergs
Störst i västerut på "gula rymmer" — Årens inkomst i Norge.

ABBY

Frankrktk.

Ryktena om förändringen uti Ministeren hafva sedan några dagar åter börjat. Enligt dessa skulle Casimir Ferier förbi ifv'n Conselj-Preklent men, i stället för Inrikes emottaga Utrikes portefeuilien, Grefve Sebastiani skulle efterträda Marskalk Soult i **Krigsdepartemenlet**, **hvilken seduare** vid Lyonska oroligheterna förfarit med mera mildhet än **Cönselj-Preudentea** velat, Tvenne nya ledamöter uppgifvas skola inträda i Ministeren, nemligen Hr Odilon Barrot för Inrikes- och **Hl' Dupin** för **Justitiae-cleparteraentet**. Nuvarande **Justiliaj-Ministern** Hr Barthe, lär, enligt **någ'a**, helt och hållet afgå, **ocb**, enligt andra, erhålla **portefeniillen** för allmänna undervisningen, hvars nuvarande innehafvare Hr **Monlalivet** i detta fall skulle öfvertaga civil-listans portefeuille. FinansMinistern Louis och MarinMinisttrn de Rigny skulle **Jaibehålla** sina platser. **Skulie** allt detta komma att äga rum, så kan Odilon Barrots gåtlika uppförande- lätteligen förklaras.

Tesseract

FRANKRYKE.

. Ryktena om förändringen uti Ministären **hafva'scdan** några dagar åter börjat. Enligt dessa skulle **Cnsimir Pericr förhliflwt Conselj**—Presiclent men, i stället för Inrikes **emoftaga** Utrikes **portefeuillem** Grefve Sebastiani skulle efterträda Marskalk **Soul!** i **Krigsdepartemeotek**, hvilken **seduarev** vid Lyonska- oroligheterna förfarit med mera mildhet än **C'onsemi**—Presidenten velat. Tvenne :nya ledamöter uppgifvas skola inträda i Ministären, **nemligeo** Hr **Odilo-u Bal'rot** för **Inrikes-c** och Hr **'Dzlpin** för Justitia—deparlementet. Nuvarande **Justitiaz**—Ministern Hr **Bm-the**, lär, enligt några, helt och **håHet** afgå, och, enligt andra, **erhååia'portefcuil[en** för **_allmä'anna** undervisningen, hvars nuvarande **innchafvare** Hr Montalivet i detta fall. **shille-** öfvertaga civil—listans **portefeuilic**. **FinansMinisten-n** Louis och **MarinMinistu—n dov Riguy skolk»** **'tihehålla** sina platser. **Skull'e** allt detta komma att äga. rum, så. **kam** Odilon Barrots **gåtliika** unnförande **låti teligen** förklaras

FRANKRYKE.
Ryktena om förändringen uti Ministeren hafva sedan några dagar åter börjat. Enligt dessa skulle Casimir Perier förblifva Conselj-President men, i stället för Inrikes emottaga Utrikes portefeuilien. Grefve Sebastiani skulle efterträda Marskalk Soult i Krigsdepartementet, hvilken seduare vid Lyonska oroligheterna förfarit med mera mildhet än Conselj-Presidenten velat. Tvenne nya ledamöter uppgifvas skola inträda i Ministeren, nemligen Hr Odilon Barrot för Inrikes- och Hr Dupin för Justitia-departementet. Nuvarande Justitia-Ministern Hr Barthe, lär, enligt några, helt och hållet afgå, och, enligt andra, erhålla portefeuilien för allmänna undervisningen, hvars nuvarande innehafvare Hr Montalivet i detta fall skulle öfvertaga civil-listans portefeuille. FinansMinistern Louis och MarinMinistern de Rigny skulle bibehålla sina platser. Skulle allt detta komma att äga rum, så kan Odilon Barrots gåtlika uppförande lätteligen förklaras.

Step	Rule	Sequence/consequence and choice of word
1.	Both words are equal = ABBYY	If ABBYY's suggestion and Tesseract's are equal: ABBYY's word is selected. [IF NOT: STEP 2]
2.	ABBY's suggestion is blank = no word is given	If ABBYY's suggestion is blank no word is given. [IF NOT: STEP 3]
3.	Both words are equal after the removal of "noise characters" = ABBYY	If ABBYY's suggestion and Tesseract's suggestion are equal after the removal of one "noise character" in the beginning and end of the given word: ABBYY's word is selected. [IF NOT: STEP 4]
4.	The word is found in ABBYY's dictionary = ABBYY	If ABBYY's suggestion is found in ABBYY's dictionary or in a customised dictionary: ABBYY's word is selected. [IF NOT: STEP 5]
5.	ABBY's suggestion is a numeral = ABBYY	If ABBYY's suggestion is a numeral that numeral is selected. [IF NOT: STEP 6]
6.	Tesseract's suggestion is blank = ABBYY	If Tesseract's suggestion is blank: ABBYY's word is selected. [IF NOT: STEP 7]
7.	If Tesseract's suggestion is a single character = ABBYY	If Tesseract's suggestion is a single character: ABBYY's word is selected. [IF NOT: STEP 8]
8.	If Tesseract's suggestion is found in Tesseract's dictionary or in a customised dictionary = Tesseract	If Tesseract's suggestion is found in Tesseract's dictionary or in a customised dictionary: Tesseract's word is selected. [IF NOT: STEP 9]
9.	If Tesseract's suggestion is a numeral = Tesseract	If Tesseract's suggestion is a numeral: Tesseract's numeral is selected. [IF NOT: STEP 10]
10.	ABBY word = ABBYY	If none of the previous steps has rendered a word: ABBYY's word is selected.

	Without word lists		With four word lists	
OCR engine	CAR (%)	WAR (%)	CAR (%)	WAR (%)
Abbyy	91.54	83.39	91.45	83.6
Tesseract	91.39	87.37	91.39	85.53
Abbyy-Tesseract	91.64	84.21	91.5	83.9

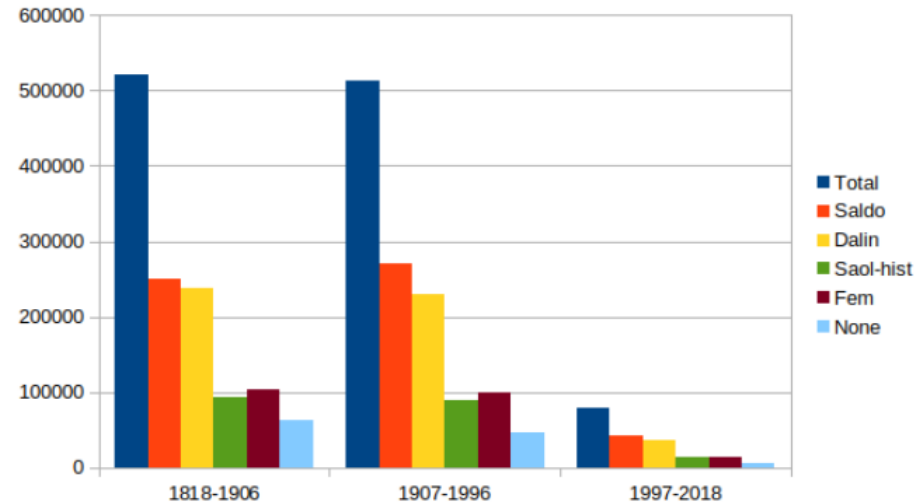
Table 1: Evaluation results of 400 pages for the whole time period 1818-2018, run with three types of OCR engine setups. Without word lists (our baseline) on the left and with all word lists on the right.

	1818-1906		1907-1996		1997-2018	
OCR engine	CAR (%)	WAR (%)	CAR (%)	WAR (%)	CAR (%)	WAR (%)
Abbyy baseline	87.38	71.47	94.22	92.11	93.34	92.58
Tesseract baseline	88.14	81.26	92.48	89.93	92.84	91.17
Abbyy-Tesseract baseline	87.46	72.66	93.36	90.05	93.43	90.11
Abbyy Dalin	87.31	71.87	94.05	91.98	93.31	92.69
Tesseract Dalin	87.86	76.74	93.32	91.44	92.55	91.07
Abbyy-Tesseract Dalin	87.32	72.49	93.13	89.7	93.34	89.92
Abbyy Saldo	87.19	71.33	93.02	89.15	93.38	89.72
Tesseract Saldo	87.89	76.73	92.46	89.39	92.95	90.21
Abbyy-Tesseract Saldo	87.18	71.96	93.05	89.42	93.27	89.78
Abbyy Saol-hist	87.37	71.96	93.13	89.44	93.47	89.83
Tesseract Saol-hist	87.84	76.75	92.46	89.38	92.95	90.19
Abbyy-Tesseract Saol-hist	87.41	72.73	93.19	89.82	93.42	90.03
Abbyy Fem	87.1	70.81	92.81	88.62	93.15	89.2
Tesseract Fem	87.88	76.72	92.45	89.37	92.96	90.2
Abbyy-Tesseract Fem	87.19	71.61	92.89	89.03	93.09	89.4
Abbyy All	87.36	71.88	93.09	89.52	93.4	89.76
Tesseract All	88.03	77.09	92.57	89.41	92.97	90.16
Abbyy-Tesseract All	87.41	72.35	93.13	89.75	93.34	89.82

Table 2: Evaluation results of 400 pages divided into three time periods. First run without external word lists (baseline), four runs with external word lists, one run for each word list and last run with all four word lists. Each run was performed with three engines. The results marked in bold highlight successful runs that outperform the baseline.

Referensmaterialet: Hur mycket täcker våra lexikon?

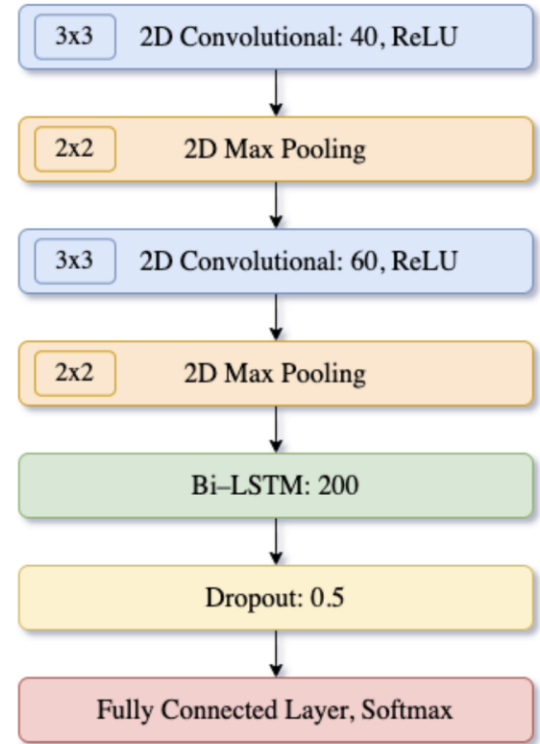
- Saldo + Saldos morfologi: 1,7M ord
- Dalin + Dalins morfologi: 509k ord
- SAOLhist: 128k ord
- Fem (namnenheter): 311k ord



Clamari: CNN-LSTM model

- 30 tidningar 1818-1848
totalt 60 sidor
- Korrekthet varierar från 56% till 97%
baseline 82.37% CAR

Dataset	GT lines	Words	Characters	B:A
Training	6 742	53 963	338 701	75% : 25%
<i>training set</i>	5 400	N/A	N/A	
<i>validation set</i>	1 349	N/A	N/A	
Testing	1 671	13 478	84 713	75% : 25%
Total	8 413	67 441	423 414	



Experiment

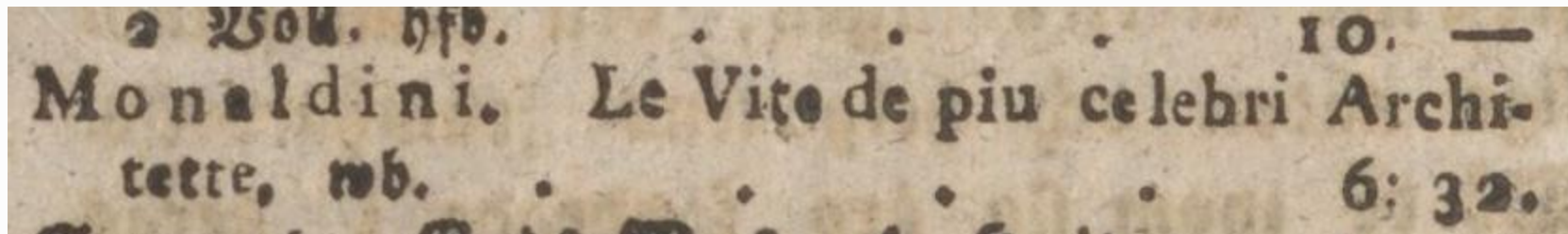
	Model(s)	Dataset	CAR (%)	CER (%)	Voted
		Test set			
(1)	M1		95.55	4.45	No
(2)	M2		96.06	3.94	No
(3)	M3		96.44	3.56	No
(4)	M4		97.43	2.57	Yes
(5)	M4	50-fraktur	96.84	3.16	Yes
(6)	M4	50-antiqua	96.48	3.52	Yes

Skelby and Dannélls (2021): OCR Processing of Swedish Historical Newspapers Using Deep Hybrid CNN-LSTM Networks. Proceedings of Recent Advances in Natural Language Processing (RANLP)

Resultat

NEWSPAPER EDITION	BASELINE	BEST	
	CAR (%)	CAR (%)	+/- (%)
STOCKHOLMSPOSTEN 1818-09-23	61.08	96.87	+35.79
GÖTHEBORGS ALLEHANDA 1819-05-26	56.65	94.79	+38.14
CARLSCRONAS WEKOBLAD 1820-05-10	78.59	94.89	+16.30
GÖTHEBORGSKA NYHETER 1821-10-27	80.84	94.95	+14.11
GÖTHEBORGS TIDNINGAR 1822-12-10	70.28	96.46	+26.18
DAGLIGT ALLEHANDA 1823-02-20	83.56	98.35	+14.79
WEXJÖBLADET 1824-08-21	80.55	98.85	+18.30
POST- OCH INRIKES TID. 1825-06-16	81.03	96.18	+15.15
STOCKHOLMS DAGBLAD 1826-11-27	78.76	96.03	+17.27
AFTONBLADET 1827-07-02	68.09	93.20	+25.11
HELSINGBORGSPOSTEN 1828-01-29	84.80	98.65	+13.85
DAGLIGT ALLEHANDA 1829-04-28	68.69	98.44	+29.75
NORRKÖPINGS TIDNINGAR 1830-03-30	81.97	98.44	+16.47
POST- OCH INRIKES TID. 1831-12-16	77.21	97.05	+19.84
GÖTEBORGS HAND. OCH SJÖ. 1832-10-15	92.15	97.92	+5.77
GÖTHEBORGS ALLEHANDA 1833-08-30	84.21	98.97	+14.76
MALMÖ ALLEHANDA 1834-03-12	87.41	98.34	+10.93
WEXJÖBLADET 1835-09-18	87.88	98.34	+10.46
POST- OCH INRIKES TID. 1836-12-08	96.40	98.46	+2.06
GEFLEBORGS LÄNS TIDNING 1837-02-01	82.15	98.49	+16.34
FREJA 1838-05-18	95.48	98.08	+3.32
CARLSCRONAS WEKOBLAD 1839-09-25	74.54	97.74	+23.2
SVENSKA BIET 1840-12-16	97.08	97.24	+0.16
NAJADEN 1841-09-03	95.36	95.78	+0.58
NORRLANDSPOSTEN 1842-01-21	85.97	96.95	+10.96
GÖTEBORGS HAND. OCH SJÖ. 1843-06-03	91.10	98.02	+6.93
WERMLANDSTIDNINGEN 1844-06-05	85.08	98.04	+12.96
JÖNKÖPINGSBLADET 1845-05-03	86.0	97.74	+11.74
NERIKES ALLEHANDA 1846-03-18	85.29	96.76	+11.44
GÖTHEBORGSKA NYHETER 1847-07-24	86.00	98.65	+12.65
SNÄLLPOSTEN 1848-05-22	89.41	99.73	+10.32
Average	82.37%	97.43%	+15.06

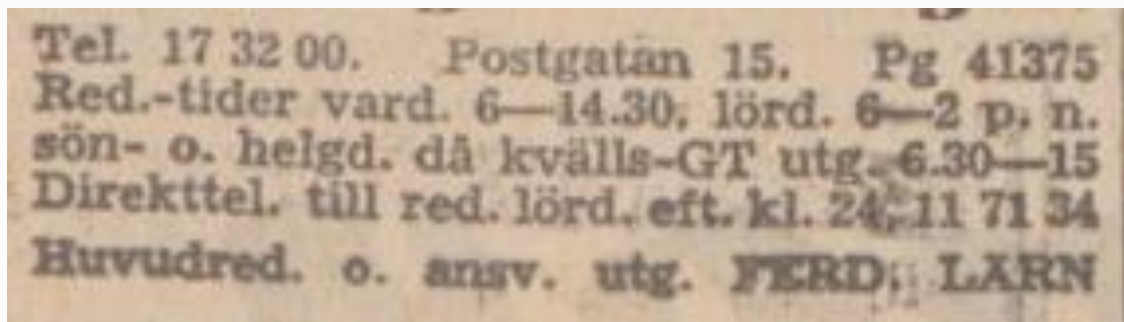
Felidentifisering av siffror



Montldini. Le Vite de piu celebri Archi-
tctte, »b. Qté JS

Montldini. Le Vite de piu celebri Archi-
tette, nr. 6;32.

Felidentifiering av förkortningar och akronymer



Tel. 17 32 00. Postgatan 15. Pg 41375
Red.-tider vard. 6—14.30, lörd. fr-2 På
Nn. o- helgd. då lörd eft kl SåRIL 71 Så
Direkttel. till red. lörd. eft. kl. 24511 7134
Huvudred, o. arav. utg. FKRD; LARN

bib20791221_19630101_149530_1_0002_003

Felidentifiering av versaler i början av en mening

HEDAKTION: Drottninggat. 55.'

Ht 57 92. At. 228 00.

Oskar Eklund, huvudredaktör och ansvarig utgivare.

Jalmar Furuskog, Filipstad.

Ri. 125.

Sigfrid Borgström, redaktionssekreterare. Träffas dagligen 12—1.

EXPEDITION: Drottninggat. 55.

At 2800 KRt: 5792



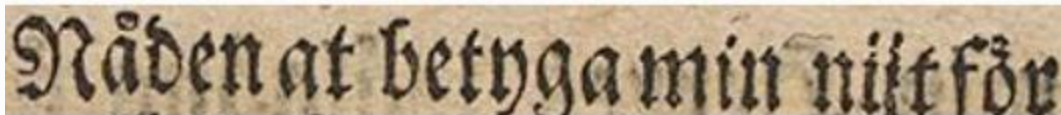
Några insikter

Ibland är resultaten motsägelsefulla, t.ex. kan lägre upplösningar eller komprimerade filer ge bättre resultat fastän det borde vara tvärt om.

Olika fel i system gör det svårt att på ett logiskt sätt arbeta fram metoder för bättre OCR.

En faktor som ytterligare försvårar arbetet är att det är svårt att veta hur verktygen arbetar med OCR internt.

Post-OCR korrigerering



Råden at betyga min niit för

Abbyy

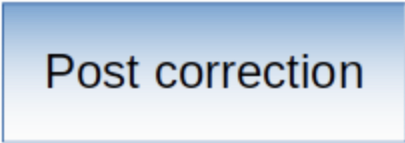
SRabett afbeipga min nt/tför

Tesseract

Nåjdenatjbetygasmtissnistföni

Ocropus

Rådenat betygamin niitför



Post correction

Råden at betyga min niit för

Post-OCR med transformer model ByT5

Input: Den i HandelstidniDgens g&rdagsnmmmer omtalade hvalfisker

Word: Den i <00V> <00V> omtalade <00V>

Sub-word: Den i Handels tid <00V> gens <00V> dags <00V> mer

Character: D e n i H a n d e l s t i d n i D g e n s

Löfgren and Dannélls (2024): Post-OCR Correction of Digitized Swedish Newspapers with ByT5. Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL)

Större material

Dataset	Partition	Time period	Characters
Newspapers	<i>Tesseract</i>	1818–2018	6 956 748
	<i>Abbyy Finereader</i>		6 928 171
Literature		1836–2001	7 266 866
Swedish fraktur		1626–1816	282 432
Then swänska Argus		1732–1734	259 468



Evalueringresultat

Period	CER (%)				WER (%)			
	BL	M1	M2	M3	BL	M1	M2	M3
1818–1859	8.39	4.39	4.63	4.30	32.46	15.34	15.80	15.54
1860–1899	4.04	2.29	2.61	2.38	16.51	8.06	8.63	8.22
1900–1939	2.60	2.01	1.97	1.92	11.24	7.03	7.08	6.99
1940–1979	1.46	1.39	1.49	1.29	6.45	4.43	4.54	4.39
1980–2018	0.83	0.67	0.75	0.73	3.74	2.67	2.67	2.67
1818–2018	3.20	2.06	2.20	2.04	13.21	7.17	7.42	7.23

OCR Cn tifiig och roff Yngling, förfebb meb
fäfra od) gobä betpg cm fitt uppföran:
be, fan erbåda fonbition i 3ernf,,(,mlim'l'ft!
fimarcht närmare unberrättelfe crfmdCb
uti >r Srobbfrnmhanblaren C. SI. UBcberé
95ob mib lida 9iögatan.

Output En tillig och rost Yngling, försedd med
säkra och goda betyg om sitt uppförande,
kan erbåda kondition i Jernf. Malmölks
simarcht närmare underrättelse erfwärd
uti _ Stockholmhandlaren C. M. Weders
Bod wid lida Sjögatan.

GT En liflig och rask Yngling, försedd med
säkra och goda betyg om sitt uppförande,
kan erhålla kondition i Jernkramhandel;
hwarom närmare underrättelse erhålles
uti Hr Kryddkramhandlaren E. A. Webers
Bod wid lilla Nygatan.

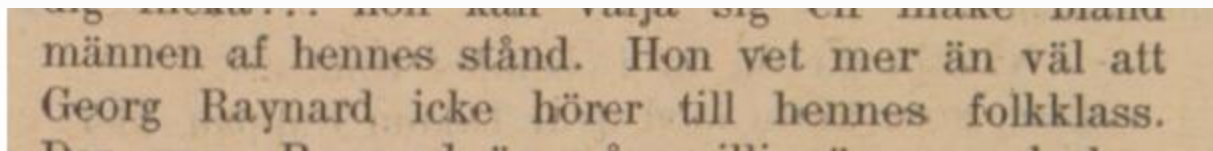
Fel analys på grund av radbrytning

coeur. Den andra har emellertid varit mera slug och vi hafva därför nu såsom alltid blifvit slagne till slantar.

OCR text - Den andra har emellertid varit mera slug och vi hafva därför nusåsom alltid blifvit slagne till slantar.

OCR correction - Den andra har emellertid varit mera slug och vi hafva därför nusåsom alltid blifvit slagne till slantar.

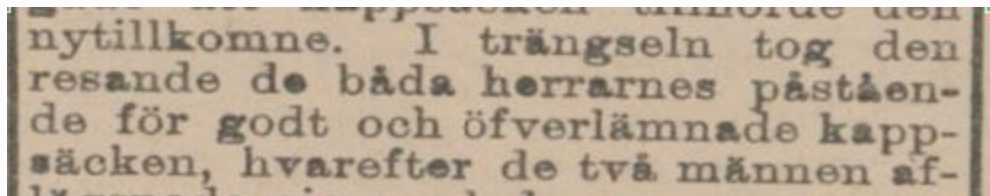
Korrekt analys



OCR text - Hon vet mer än väl attGeorg Raynard icke hör till hennes folkklass.

OCR correction - Hon vet mer än väl att Georg Raynard icke hör till hennes folkklass.

Blandade resultat



OCR text - I trängseln tog **denresande** **ds** **bida** **hsrrarnes** påstående för godt och öfverlämnade kapptäcken,

OCR correction - I trängseln tog **denresande** **de** **bida** **herrarnes** påstående för godt och öfverlämnade kapptäcken,

Title: — Amerikansk annonsör i Stockholm. Enbrefkrifware till Norrköpings tidningar berättar:En

Close Document

Corpus name ID: KALMAR NEWSPAPER ORIGINAL 1872

Document size: 199 tokens

Year: 1872

 Most common nouns:
 rdr. (2)
 pris (2)
 annonsör (1)
 Enbrefkrifware (1)
 berättar:"En (1)
 handlande (1)
 synd (1)
 läsare (1)

 uppbyggelseicke (1)
 infriptioner

 Most common names:
 U. F. Enström (3)
 Herr Enström (2)
 Stockholm (1)

Document view

Statistics view

Explore

word attributes ▾

no choice ▾

Q Search the current document

1 — Amerikansk annonsör i Stockholm. **Enbrefkrifware** till "Norrköpings tidningar" berättar: "En handlande U. F. Enström **å Westerlånggatan** i Stockholm har lärt sig annonsera och **flyta**. **Det** är verkligen synd att wi till läsarens **å uppbyggelseicke** antecknat de **infriptioner**, med hwilka de papperslappar äro försedda, som äro satta **å de warorhan** **S** utlagt i sitt fönster. Wi wilja dock ur minnetmeddela det ungefärliga innehållet. **Å** en **fjortäläses**: "För att **å gå** ärade kunders **önffningar** tillmötes har undertecknad ökat längden i sina **fjortor**." **Å** en näsdukswäf: "Otroligt, **Å** men sannt! Etddussin näsdukar **fås** för 1 rdr. Hwad tycks? **Åen** halsduk: "Kostar bara rdr. **F—u så billigt!**" **ÅH** ett paraply: "Tak **öfwer hufwudet** för 3 **rdr**." **Å** en portmonnä: "5 rum och kök för 75 **öre**." **Oswcrst** i fönstret prunka **Å** orden: "Hwilka waror **—hwilka pris!**" Herr Enström **mätte** wara en mycket förekommande och framför allt **menniflowänligman**, åtminstone att döma af hans yttranden. **Det** är t. ex. intet owanligt att höra honom säga: "**Ja, mitt herrflap**, prisen förefalla er sannolikt **otroligtbilliga**, men jag försäkra er att det är sannt. **Ja, jag** går min ruin, min undergång till mötes förde billiga prisens **full**, men publiken är beläten —och då är jag nöjd." Hwilken **sjelfuppoffring!** **Mannen** gör sig förtjent af den **flönaste mcdborgarekrona**."

2

3

1 — Amerikansk annonsör i Stockholm. **Enbrefkrifware** till "Norrköpings tidningar" berättar: "En handlande U. F. Enström **S Westerlånggatan** i Stockholm har lärt sig annonsera och **skylda**. **Det** är verkligen synd att wi till läsarens **å uppbyggelse icke** antecknat de **inskrptioner**, med hwilka de papperslappar äro försedda, som äro satta **S de waror han** **å** utlagt i sitt fönster. Wi wilja dock ur minnetmeddela det ungefärliga innehållet. **S en skjortäläses**: "För att **å gå** ärade kunders **önskningar** tillmötes har undertecknad ökat längden i sina **skjortor**." **S en** näsdukswäf: "Otroligt, **Å** men sannt! Etddussin näsdukar **fås** för 1 rdr. Hwad tycks? **en** halsduk: "Kostar bara 1 rdr. **F—n så billigt!**" **Å** ett paraply: "Tak **öfwer hufwudet** för 3 **rdr**." en portmonnä: "5 rum och kök för 75 **öre**." **Öfwerst** i fönstret prunka **Å** orden: "Hwilka waror **= hwilka pris!**" Herr Enström **mätte** wara en mycket förekommande och framför allt **menniskowänligman**, åtminstone att döma af hans yttranden. **Det** är t. ex. intet owanligt att höra honom säga: "**Ja, mitt herrskap**, prisen förefalla er sannolikt **otroligt billiga**, men jag försäkra er att det är sannt. **Ja, jag** går min ruin, min undergång till mötes förde billiga prisens **kull**, men publiken är beläten —och då är jag nöjd." Hwilken **Mannen sjelf uppoffring**. gör sig förtjent af den **skönaste medborgarekrona**."

2

3

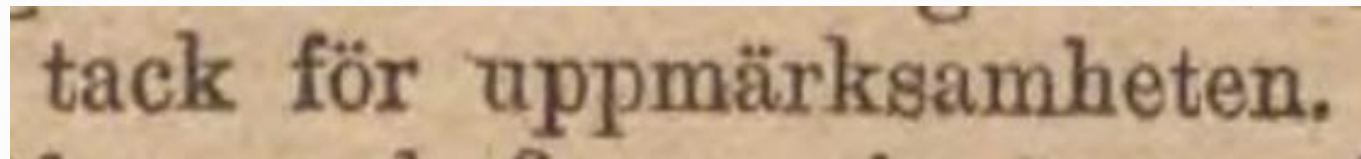
Slutord och nästa steg

Den nya OCR-modellen ger hög igenkänningsnoggrannhet för äldre material med sämre tryckkvalité.

Modellen implementerades i Strix och Sprav – SBX:s och kommer att göras tillgänglig till hösten.

Genomföra en större evaluering på mångfaldigt material.

Träna post-OCR modellen med en större ByT5-model genom att ta hänsyn till kontexten och längre ordsekvenser.



tack för uppmärksamheten.

Dana Dannélls <dana.dannells@svenska.gu.se>

<https://spraakbanken.gu.se/>