

# Ammaseus Horisont - Dokumentation

Ett projekt för att lära artificiell intelligens tolka Karin Boyes lyrik.

## Innehållsförteckning

<b>1.</b>	<b>Introduktion</b>	2
<b>2.</b>	<b>Kort om artificiella neurala nätverk</b>	2
2.1	<i>Feedforward-nätverk</i>	3
2.2	<i>Feedback-nätverk</i>	4
2.3	<i>Long Short Term Memory-nätverk</i>	5
<b>3.</b>	<b>Träning av LSTM-nätverk</b>	5
3.1	<i>Standardmodeller</i>	5
3.1.1	<i>Ordbaserad LSTM</i>	6
3.1.2	<i>Bokstavsbaserad LSTM</i>	7
3.2	<i>Expanderade modeller</i>	8
3.2.1	<i>Open AI:s GPT-2</i>	9
<b>4.</b>	<b>Textredigering</b>	11
<b>5.</b>	<b>Resurser</b>	12
<b>6.</b>	<b>Appendix – träningsdata</b>	12

## 1. Introduktion

*Stor är människans strävan,  
stora de mål hon satt –  
men större är människan själv  
med rötter i alltets natt.*

Sedan Karin Boye skrev dessa ord för snart hundra år sedan har mycket hänt, men aldrig har de varit mer relevanta än idag. Människans strävan för att skapa intelligenta maskiner har inte bara lett till AI-program som slagit människan i Schack och Go, spel som i urminnes tider symboliserat intelligens, utan de senaste åren har AI även skapat konstverk och komponerat musik på en nivå som gjort det svårt till och med för experter att skilja verken från mänskliga kreationer.

Det här projektet tar sin utgångspunkt i frågan om det är möjligt för en AI, eller mer specifikt, ett så kallat djupt neuralt nätverk, att lära sig skriva poesi i stil med en specifik mänsklig diktare. Och vem bättre än Karin Boye, en av våra största poeter, vars diktning genomgående brottas med de stora existentiella frågorna om vad det innebär att vara människa.

*Några tekniska noteringar.*

För implementering av neurala nätverk i detta projekt har främst programbiblioteket Tensorflow för Python använts. Ett programbibliotek utvecklat av Google som gjort det relativt enkelt att skapa och träna artificiella neurala nätverk.

## 2. Kort om artificiella neurala nätverk

*Nedan beskrivning av artificiella neurala nätverk är mycket översiktligt och utelämnar detaljer. Den ges endast som en kort bakgrund till vad för slags modeller som använts i projektet.*

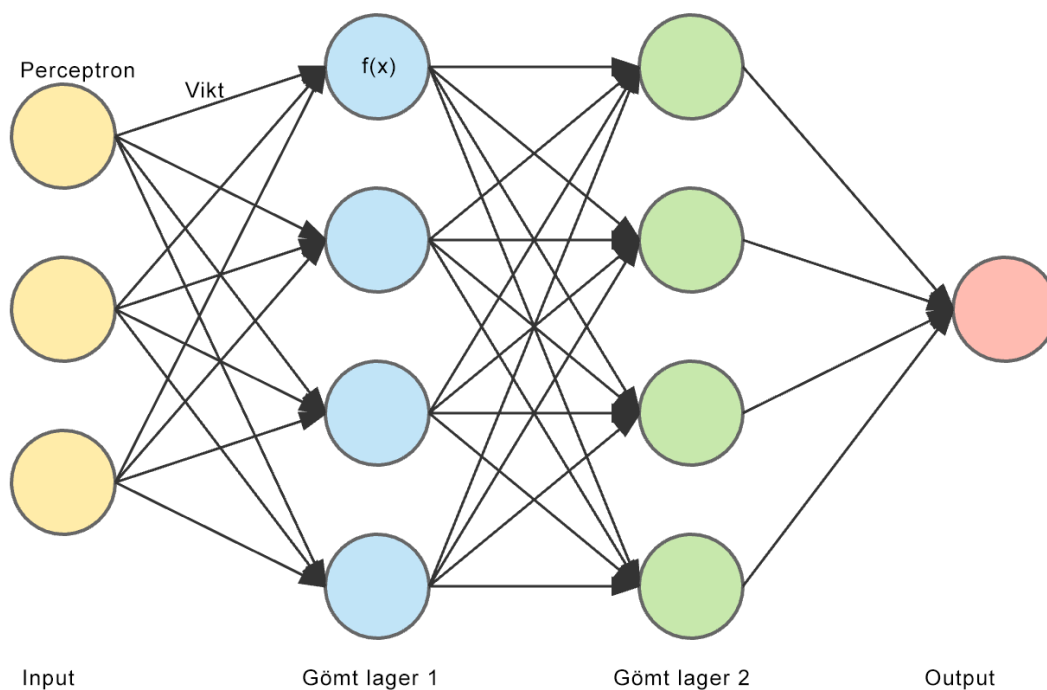
Ett neuralt nätverk är, mycket enkelt beskrivet, ett nätverk, bestående av så kallade perceptroner (artificiella neuroner), som är kopplade till varandra på ett vis som är inspirerat av den biologiska hjärnans neuroner och synapser.

Varje perceptron tar emot data (input) och sänder ut data (output). Varje koppling mellan två perceptroner har en så kallad "vikt" vilket är en variabel som avgör vad inputen skall multipliceras med innan den tas emot av perceptronen. I perceptronen behandlas sedan detta värde av en tröskelfunktion, även kallad aktiveringsfunktion, som avgör om och i så fall vilket output som perceptronen skall ge. Denna funktion kan se ut på många olika sätt, vilket inte kommer att beskrivas i den här dokumentationen.

I ett djupt neuralt nätverk är perceptronerna ordnade i flera lager, där perceptronerna i det ena lagret kopplar till perceptronerna i nästa lager och så vidare. De nätverk som diskuteras nedan är så kallade "fully connected", vilket innebär att samtliga perceptroner i ett lager kopplar till samtliga perceptroner i nästa lager.

### 2.1 Feedforward-nätverk

I figuren nedan visas ett enkelt *feedforward*-nätverk, med totalt fyra lager; ett input lager och ett outputlager samt två så kallade gömda lager däremellan. Varje linje i figuren motsvarar en koppling mellan två perceptroner och har således en vikt. Vidare har varje perceptron en aktiveringsfunktion,  $f(x)$ , som avgöra vilken output perceptronen skall ge. I djupa neurala nätverk finns ofta betydligt fler än två gömda lager och ofta betydligt fler perceptroner i varje lager än i den förenklade figuren nedan:

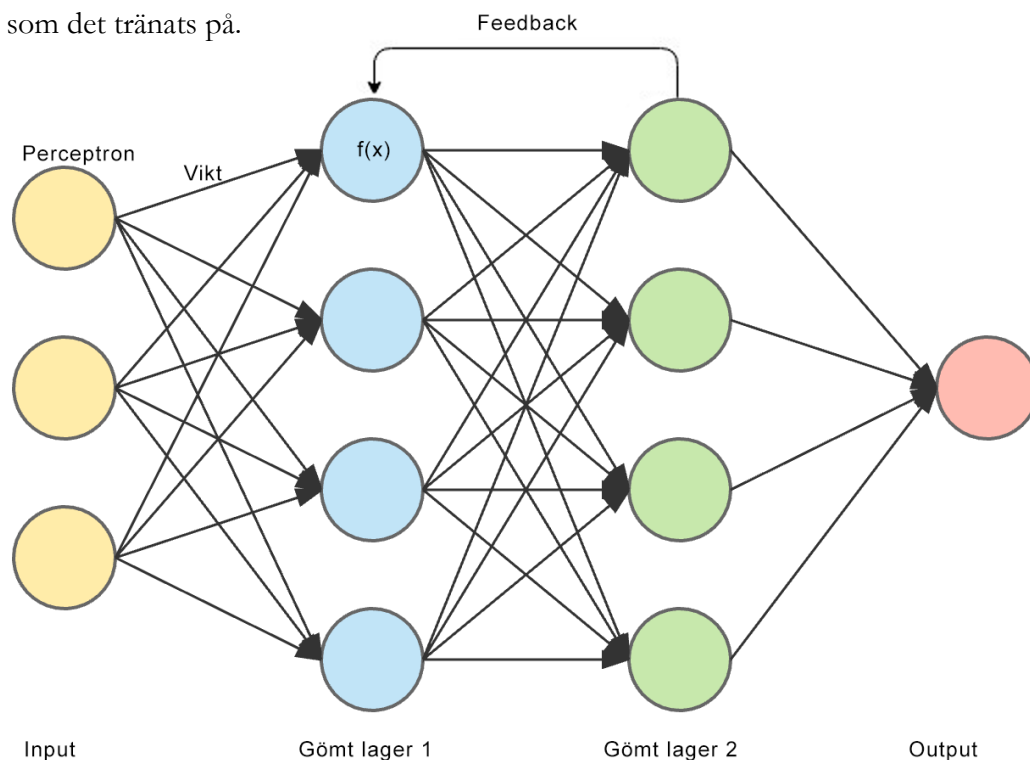


För att träna ett *feedforward*-nätverk ges träningsdata till inputlagret och detta data behandlas sedan via vikterna och perceptronernas aktiveringsfunktioner genom nätverket, för att sedan ges som output i outputlagret. Denna output jämförs sedan med "facit" för träningsdatat, dvs. vad outputen egentligen skulle ha varit för varje given input, och därefter justeras vikterna mellan perceptronerna i flera cykler tills outputen närmar sig träningsdatats facit. Principen är att undersöka i vilken riktning felet är på väg (dvs. blir felet större eller mindre om vikten ökar eller minskar). Denna optimeringen av nätverket under träningen kan ske på många olika sätt, vilket inte kommer att gås in vidare på i den här dokumentationen.

Ett *feedforward*-nätverk kan utifrån given data ge en output som representerar det input som blev givet. Sådana nätverk är bra på bland annat klassificeringsproblem, där det t.ex. handlar om att avgöra ifall en given input, t.ex. en bild, tillhör en given klass, till exempel en katt eller en bil. Däremot fungerar det inte bra för att förutse ett kommande tillstånd som beror av ett tidigare tillstånd, vilket är fallet för sekventiellt data så som text, där nästa bokstav eller ord i en följd beror på vilka bokstäver eller ord som kommit innan. För detta behövs en annan typ av nätverk som inkluderar någon form av feedback.

## 2.2 Feedback-nätverk

För sekventiell data har en viss typ av layout visat sig speciellt effektivt. Ett så kallat Recurrent Neural Network (RNN), som är en speciell klass av artificiellt neuralt nätverk. I ett RNN behålls outputen från föregående tillstånd och används som input igen, tillsammans med eventuellt nya data. Det vill säga, nätverket fortsätter producera output med hjälp av föregående output som input. På detta sätt kan sannolikheten för att en viss bokstav skall följa efter en annan uppskattas av nätverket utifrån den text (träningsdata) som det tränats på.



### *2.3 Long Short Term Memory - nätverk*

Ett problem med enkla RNN som det ovan, är att det har svårt att minnas saker över tid. Föregående bokstav eller ord i texten beror mest på den bokstav eller det ord som kommit precis före, men när informationen som är relevant ligger långt bak i tiden så misslyckas ett sådant RNN. Den har ett bra korttidsminne men ett dåligt långtidsminne och för textgenerering är det inte bara det föregående ordet eller bokstaven i följd som är viktig, utan även ord eller bokstäver som förekommit tidigare i texten.

För att komma till rätta med detta har så kallade Long Short Term Memory-nätverk utvecklats (RNN LSTM). En perceptron i ett LSTM nätverk har till skillnad från en perceptron i ett vanligt nätverk, ett eget långtidsminne. Utan att gå in på detaljer, så består en LSTM perceptron typiskt av tre stycken "gates", en som tar emot input, en som avgör vilken output som skall ges ut, och en som avgör vad som skall minnas och vad som skall glömmas. Denna design har visat sig effektiv för att hantera sekventiell data, så som text.

## **3. Träning av LSTM-nätverk**

En viktig aspekt när det gäller att träna ett neuralt nätverk är träningsdatat. Generellt krävs det stora mängder data och att detta data får köras ett stort antal cykler genom nätverket. Tränas nätverket för kort tid, blir resultatet slumpmässigt och texten osammanhängande. Tränas däremot nätverket på en för liten datamängd eller för länge i relation till nätverkets storlek, så är risken stor att nätverket istället för att lära sig generalisera, dvs lära sig en "stil" på texten den tränas på, börjar memorera texten. Detta kallas överträning.

De första försöken i det här projektet gjordes med relativt små standardnätverk och träningsdata som innehöll enbart Karin Boyes dikter. Här nedan visas några resultat från dessa tidiga försök.

### *3.1. Standardmodeller*

En grundläggande fråga för textgenerering är på vilken nivå genereringen ska ske. Om det ska vara på bokstavs nivå eller på ordnivå. Det vill säga, ska inputen vara ett ord och ska nätverkets uppgift vara att gissa nästa ord i följd, eller skall det hela ske på bokstavs nivå. Varje modell har sina för- och nackdelar. Om det sker på bokstavs nivå så krävs betydligt fler träningscykler, nätverket måste förutom att lära sig sätt ihop ord också

lära sig att skapa ord och att stava. Men en av fördelarna är att det för varje given input finns ett mer begränsat antal möjliga output.

Fördelen med ett nätverk som istället arbetar på ordnivå är att det inte måste lära sig stava och går betydligt snabbare att träna. Däremot kommer ett sådant nätverk aldrig generera några andra ord än de som finns i träningsdatat och innan nätverket kan börja tränas måste även alla ord i träningsdatat läsas in i en lista som sedan används som referens för alla möjliga output.

Vad som är bäst beror på omständigheterna men generellt anses nätverk som arbetar på bokstavs nivå prestera bättre, även om de kräver betydligt mer träning.

Båda metoderna utvärderas här nedan.

### *3.1.1 Ordbaserat LSTM*

De första försöken gjordes med ett relativt standardiserat LSTM-nätverk för textgenerering på ordnivå. Nätverket hade fyra gömda lager med 100 perceptroner i varje lager.

Efter 200 träningscykler genererades text som verkade relativt slumpmässig:

```
och väntar har väntar har  
slutar väntar klamrar sig  
ur varför men väntar
```

Men redan efter 400 cykler började texten påminna om poesi:

```
Skal men varför har du gömt dig så väl för mig  
det mal och mal  
jag vet det jag minns  
det en enda sak där jag har dömt
```

Efter en kontroll mot träningsdatat visade det sig emellertid att modellen redan hade börjat plagiera. Här är originaltexten från Karin Boyes egna verk:

```
Där var ditt skal.  
Men varför har du gömt dig så väl för mig?  
Det mal och mal.
```

```
Jag vet det. Jag minns det: en enda sak,  
där jag har dömt -
```

Det var tydligt att nätverket övertränats, och det visade sig sedan svårt att hitta ett mellanläge där texten blev bra utan att vara plagiat.

### 3.1.2 Bokstavsbaserad LSTM

Ett relativt standardiserat LSTM-nätverk på bokstavnivå med tre gömda lager och 256 perceptioner i varje lager, utvärderades.

Här är några tidiga resultat efter att ha tränat nätverket:

10 träningscykler:

```
si lan van sag san san san
```

100 cykler:

```
an digan saka  
han salter sicla  
gör dn samne skälk  
vilar sid soanka  
av oin och dedar
```

500 cykler.

```
min döda som fot staras  
av el oor anir  
sår met
```

1000 cykler

```
det mörkrädda att  
det vara  
lampa stjärneblick som  
livet är skål
```

Därefter stagnerade inlärningen och texterna blev bara marginellt bättre och genererade ganska liten variation.

För dessa standardnätverk gick resultaten alltså ganska snabbt från oläsliga, till slumpmässiga, och det stod klart att nätverken måste göras betydligt större för att generera unik text som kändes genuin men utan att vara rena plagiat.

### 3.2 Expanderade modeller

Nästa steg blev att utvärdera större bokstavsbaserad LSTM-nätverk. När storleken blev upp emot 5-6 gömda lager med 512 perceptroner i varje lager, blev nätverken för stora för att kunna tränas på en vanlig hemmadator och måste istället tränas på GPU:er (Graphic Processing Units), vilket gjordes via Googles cloud computing tjänst.

För dessa större nätverk blev överträning ett ännu större problem. Det gick relativt snabbt för dem att memorera hela textstycken. För att undvika överträning måste träningen delas upp i flera steg, där nätverket först tränades längre perioder på stora mängder generiskt text, för att sedan successivt tränas på mindre och mer specifika datamängder. Bäst blev resultaten när träningen delades upp i fyra steg med fyra olika datamängder (lista över vad dessa datamängder innehåller finns i appendix.)

#### Steg 1

Nätverket tränas på *datamängd 1*, vilket är en stor generisk textmängd på svenska, tagna från svenska wikipedia-sidor, totalt cirka 10 GB text.

#### Steg 2

I nästa steg tränas nätverket på en mindre datamängd, (*datamängd 2*) med utvalda litterära verk från Projekt Runeberg. Totalt cirka 90 MB text. Detta för att få nätverket att generera svensk text med litterära kvalitéer.

#### Steg 3

Nätverket tränas sedan på *datamängd 3*, en delmängd ur *datamängd 2* som enbart innehåller utvalda stycken lyrik, men av fler diktare än Karin Boye.

#### Steg 4

I sista steget används *datamängd 4* som är en delmängd ur *datamängd 3*, enbart innehållande Karin Boyes diktning.



I rutinen ovan är varje ny datamängd i hierarkin mindre och mer specifik än föregående och nätverket tränas allt färre cykler per ny datamängd för att undvika överträning.

Under arbetet med att utvärdera olika layouter och träningsparametrar för nätverken annonserad Open AI att de släppte källkoden till sin mycket kraftiga modell för textgenerering som de kallade GPT-2. I sektionen nedan utvärderas denna modell, vilken också i slutändan blev den som presterade bäst.

### *3.2.1 Open AI:s GPT-2*

I november 2019 släppte Open AI sin GPT-2 modell. De hade tvekat att släppa den med rädsla för att modellen var för kraftfull och bland annat skulle kunna användas för att generera ett överflöd av falska nyheter, men genom att i steg släppa allt större versioner av modellen, vågade de tillslut släppa hela. Modellen blev snabbt uppmärksam för att den kunde generera långa stycken text som inte gick att skilja från mänskligt skriven text.

Tillsammans med GPT-2 modellen (versionen med 24 lager och totalt 355 miljoner parametrar) släpptes även en färdigtränad modell som hade tränats på över 40 GB engelsk text insamlade från över 8 miljoner webbsidor. Modellen genererade därför bara engelsk text, så den behövde tränas om för att lära sig svenska.

På grund av modellens storlek, mängden data den behövde tränas på och antal cykler som skulle behövas, så var det uteslutet att göra det på en vanlig hemmadator. All träning av GPT-2 har skett på en NVIDIA Tesla K80 GPU maskin via Googles cloud computing tjänst.

Efter att ha utvärderat ett antal olika träningsprotokoll med de fyra datamängderna ovan så visade det sig att resultaten blev bäst om modellen innan den fick mer specifik träning, först tränades cirka 20 000 cykler på datamängd 1 (generisk svensk text från Wikipedia), och sedan ytterligare 18 000 cykler på utvalda litterära verk (datamängd 2). De olika datamängdernas innehåll finns listade i appendix.

Här är några resultat från träningen på datamängd 2:

5 000 cykler:

Han blev nästan sällan värkande, och fällde en man, som rörde sig in i den myckenhet, att han skulle kunna vässa spelmännen vid detta rum.

18 000 cykler:

Hör! Nu säger vi så: det är en skugga, en frisk och djupt dundrande  
skugga, den är bara liksom en backe, ingen kan släcka den. Och du  
behagar säga nej?

Vid det här laget genererades mestadels mer eller mindre koherent svensk text. Därefter  
tränades modellen 2 000 cykler på datamängd 3 (som enbart innehåller lyrik) och därefter  
ytterligare 1 800 cykler på datamängd 4 (enbart Karin Boyes lyrik).

Resultaten under träningen med datamängd 4 (Karin Boyes dikter) kunde se ut som följer:

500 cykler:

Du blommar och guld, guld  
Väntar på långt innan mörker  
och stelnad känner du.

1 000 cykler:

Vacker är måsarna,  
lösta från resorna  
av himlafurstarnas ängel.

1 800 cykler:

Jag ville veta, vad som väntade  
de döda har sökt oss  
i sömn låg vi hemma.  
Ammaseus horisont  
Ammaseus Horisont

Här började dikterna se koherenta ut och dessutom vävdes vid enstaka tillfällen ordföljder  
från Karin Boyes diktning in. För att kontrollera hur mycket av den genererade texten  
som tagits direkt från Karin Boyes diktning jämfördes outputen med träningsdatat.  
Enbart i sällsynta fall kopierades hela meningar, som t.ex. i det här fallet:

Min Gud  
och min sanning  
så många gudar  
och så många sanningar  
och ändå Gud.

Där Karin Boyes original lyder:

Min Gud  
och min sanning  
såg jag  
i en sällsam stund.  
Mänskors ord  
och bud tego.

De två första raderna är alltså plagierade från Karin Boyes originaltext men därefter är den genererade texten unik och därtill knyter den tematiskt an till de första raderna. Men detta var ett undantag, och i de allra flesta fall var texterna som genererades unika, men ofta med teman som t.ex. Gud, som även frekvent återfinns i Karin Boyes egen diktning.

Intressant nog så har modellen även vid några tillfällen skapat helt egna ord, som t.ex.:  
*Ammaseus.*

Den slutgiltiga modellen, har alltså förutom att vara förtränad på det engelska språket på cirka 40 GB data, därefter tränats totalt 41 800 cykler i fyra steg med fyra olika datamängder. Den totala träningsstiden för detta uppgick till cirka 20 timmar på en NVIDIA Tesla K80 GPU-maskin.

#### 4. Textredigering

Eftersom modellen genererar text genom att ”gissa” vilken bokstav som skall vara nästa i en bokstavsföljd, så kräver den färdigtränade modellen ett ”frö” för att börja generera text. Fröet som användes här var ”varde ljus” och sedan fick helt enkelt modellen producera en textfil med cirka 30 000 tecken. Ur detta material har ett visst urval gjorts, det vill säga vissa dikter har tagits bort, till exempel då de innehållit hela meningar tagna direkt ur Karin Boyes egen diktning, eller då de varit alltför repetitiva, vilket ibland kunde var fallet. Men på det stora hela har texten tagits precis som den genererades. Ur originaldokumentet på ca 30 000 tecken, återstår i den färdiga versionen över 22 000.

I övrigt har redigeringen av texten hållits till ett minimum för att resultatet skall vara så genuint som möjligt. Uppenbart felstavade ord har rättats och bokstäver efter punkt har genomgående gjorts om till versaler. I vissa fall har radbryt lagts till för att skilja strofer och/eller dikter från varandra. Men i det stora hela är dikterna oredigerade och därtill satta i samma ordning som de genererades av modellen.

## 5. Resurser

Mer material, länkar, resultat och träningsdata samt instruktioner för hur den färdigtränade modellen kan laddas ner och testas, finns tillgängliga på projektets Github repository: <https://github.com/svenskscifi/karinboye/>

## 6. Appendix – träningsdata

### Datamängd 1

*10 GB generisk svensk text från Wikipedia*

### Datamängd 2

*Litterära verk tagna från Projekt Runeberg:*

Dan Andersson - Dan Anderssons samlade skrifter: minnesupplaga  
Dan Andersson – Dikter i Det kallas vidskepelse samt Brooklandsvägen  
Dan Andersson – Dikter  
Dan Andersson – Efterlämnade dikter  
Dan Andersson – Kolvaktarens visor  
Dan Andersson – Svarta Ballader  
Dan Andersson – Visor och ballader  
Edith Södergran – Brokiga iakttagelser  
Edith Södergran – Dikter  
Edith Södergran – Framtidens skugga  
Edith Södergran – Landet som icke är  
Edith Södergran – Rosenaltaret  
Edith Södergran – Septemberlyan  
Edith Södergran - tankar om naturen  
Gustaf Fröding – Fyra dikter  
Gustaf Fröding – Gitarr och dragharmonika  
Gustaf Fröding – Gralstänk  
Gustaf Fröding – Grillfängerier  
Gustaf Fröding – Mattoidens sånger  
Gustaf Fröding – Nya dikter  
Gustaf Fröding – Om Lifsmonader  
Gustaf Fröding – Skrifter, jubiliemsupplaga  
Gustaf Fröding – Stänk och flikar  
Erik Axel Karlfeldt – Efterskörd  
Erik Axel Karlfeldt – Erik Axel Karlfeldts dikter  
Erik Axel Karlfeldt – Flora och Bellona  
Erik Axel Karlfeldt – Flora och Pomona  
Erik Axel Karlfeldt - Fridolins lustgård och Dalmålningar på rim  
Erik Axel Karlfeldt - Fridolins lustgård och andra dikter

Erik Axel Karlfeldt – Hösttorn  
Erik Axel Karlfeldt – Vildmarks och kärleksvisor  
Karin Boye – De sju dödssynderna och andra efterlämnade dikter  
Karin Boye – För trädets skull  
Karin Boye – Gömda land  
Karin Boye – Härdarna  
Karin Boye – Moln  
Karin Boye – Kallocaïn  
Erik Johan Stagnelius – Lyriska dikter intill tiden omkring 1818  
Erik Johan Stagnelius – Strödda dikter efter tiden omkring 1818  
Viktor Rydberg - Bibelns lära om Kristus  
Viktor Rydberg - De vandrande djäknarne  
Viktor Rydberg - Den siste atenaren  
Viktor Rydberg - Det sköna och dess lagar  
Viktor Rydberg - Dikter  
Viktor Rydberg - En underbar man : Leonardo Da Vinci  
Viktor Rydberg - En vandring i Norge  
Viktor Rydberg - Johann Wolfgang von Goethe, Faust och fauststudier  
Viktor Rydberg - Filosofiska föreläsningar  
Viktor Rydberg - Fribytaren på Östersjön  
Viktor Rydberg - Fädernas gudasaga m.m.  
Viktor Rydberg, Alice Tegnér - Goternas sång  
Viktor Rydberg - Kulturhistoriska föreläsningar  
Viktor Rydberg - Lille Viggs äfventyr på julafton  
Viktor Rydberg - Medeltidens magi  
Viktor Rydberg - Romerske kejsare i marmor samt andra uppsatser i konst  
Viktor Rydberg - Singoalla  
Viktor Rydberg - Sägner, berättelser och skizzer  
Viktor Rydberg - Undersökningar i germanisk mytologi  
Viktor Rydberg - Vapensmeden (Hägringar från reformationstiden)  
Viktor Rydberg - Varia (filosofiska, historiska, språkvetenskapliga ämnen)  
Henrik Ibsen – Brand (på norska)  
Henrik Ibsen – Et dukkehjem (på norska)  
Henrik Ibsen – Henrik Ibsens samlade verk (på norska)  
Henrik Ibsen – Peer Gynt (på norska)  
Selma Lagerlöf - Anna Svärd  
Selma Lagerlöf - Antikrists mirakler  
Selma Lagerlöf - Bannlyst  
Selma Lagerlöf - Charlotte Löwensköld  
Selma Lagerlöf - Drottningar i Kungahälla  
Selma Lagerlöf - Dunungen. Lustspel i fyra akter  
Selma Lagerlöf - En herrgårdssägen  
Selma Lagerlöf - En saga om en saga  
Selma Lagerlöf - Gösta Berlings saga  
Selma Lagerlöf - Herr Arnes penningar  
Selma Lagerlöf - Jerusalem  
Selma Lagerlöf - Kejsarn av Portugallien  
Selma Lagerlöf - Kristuslegender  
Selma Lagerlöf - Körkarlen  
Selma Lagerlöf - Legender  
Selma Lagerlöf - Liljecronas hem  
Selma Lagerlöf - Löwensköldska ringen  
Selma Lagerlöf - Mårbackablomster  
Selma Lagerlöf - Mårbacka  
Selma Lagerlöf - Nils Holgerssons underbara resa genom Sverige  
Selma Lagerlöf - Osynliga länkar  
Selma Lagerlöf - Troll och människor I  
Selma Lagerlöf - Troll och människor  
Selma Lagerlöf - Ur Gösta Berlings saga. Berättelse från det gamla Värmland  
Selma Lagerlöf - Zachris Topelius. Utveckling och mognad Swedish

August Strindberg - Abu Casems tofflor  
August Strindberg - Bjälbo-Jarlen  
August Strindberg - Carl XII  
August Strindberg - Czarens kurir eller Sågfilarens hemligheter Swedish  
August Strindberg - Den starkare  
August Strindberg - Dödsdansen  
August Strindberg - Ett drömspel  
August Strindberg - Faderen. Sorgespil i tre akter  
August Strindberg - Fadren  
August Strindberg - Fagervik och Skamsund  
August Strindberg - Fritänkaren. Dramatiskt utkast  
August Strindberg - Fröken Julie  
August Strindberg - Gustav Vasa  
August Strindberg - Hermione. Sorgspel  
August Strindberg - Kammarspel  
August Strindberg - Lykke-Pers Rejse  
August Strindberg - Mäster Olof  
August Strindberg - Ordet i min makt. Läsebok för underklassen  
August Strindberg - Riksföreståndaren : skådespel i fem akter  
August Strindberg - Röda rummet  
August Strindberg - Samlade otryckta skrifter  
August Strindberg - Stora Landsvägen. Ett vandringsdrama med sju stationer  
August Strindberg - Svenska öden och äfventyr : berättelser från alla tidevarf  
August Strindberg - Svenska öden och äfventyr  
August Strindberg - Sömngångarnätter på vakna dagar. En dikt på fria vers  
August Strindberg - Till Damaskus  
August Strindberg - Tschandala. Berättelse från 1600-talet  
August Strindberg - Utopier i verkligheten  
August Strindberg - Vid högre rätt  
August Strindberg - Dikter på vers och prosa.  
August Strindberg - Sömngångarnätter på vakna dagar  
August Strindberg - Ungdomsdramer  
August Strindberg - Mäster Olof  
August Strindberg - I vårbrytningen  
August Strindberg - Kulturhistoriska studier  
August Strindberg - Röda rummet  
August Strindberg - Gamla Stockholm  
August Strindberg - Svenska folket.  
August Strindberg - Tidiga 80-talsdramer  
August Strindberg - Det nya riket  
August Strindberg - Folkungasagan. Gustav Vasa. Erik XIV  
August Strindberg - Utopier i verkligheten  
August Strindberg - Kronbruden.  
August Strindberg - Svanevit  
August Strindberg - Ett drömspel  
August Strindberg - Fagervik och Skamsund.  
August Strindberg - Ordalek och småkonst  
August Strindberg - Sagor. Ensam  
August Strindberg - Näktergalen i Wittenberg.  
August Strindberg - Kristina. Gustav III  
August Strindberg - Götiska rummen  
August Strindberg - Svarta fanor  
August Strindberg - Historiska miniatyrer  
August Strindberg - Hövdingaminnen  
August Strindberg - Taklagsöl. Syndabocken  
August Strindberg - Kammarspel  
August Strindberg - Stora landsvägen.  
August Strindberg - Abu Casems tofflor  
August Strindberg - Språkvetenskapliga studier  
August Strindberg - Tal till svenska nationen samt andra tidningsartiklar 1910-1912

August Strindberg - Berättelser, dikter, artiklar  
August Strindberg - Han och hon. En själs utvecklingshistoria 1875-76  
August Strindberg - Kaspers fet-tisdag. Påsk. Midsommar  
August Strindberg - Dödsdansen  
August Strindberg - Engelbrekt  
August Strindberg - Kronbruden.  
August Strindberg - Svanevit.  
August Strindberg - Gustav Adolf  
August Strindberg - En blå bok  
August Strindberg - Siste riddaren.  
August Strindberg - Riksföreståndaren  
August Strindberg - Bjälbo-jarlen  
August Strindberg - Öppna brev till Intima teatern  
August Strindberg - Fagervik och Skamsund.  
August Strindberg - Ordalek och småkonst  
August Strindberg - Svenska Öden och Äventyr  
August Strindberg - Giftas  
August Strindberg - Likt och olik  
August Strindberg - Tjänstekvinnans son  
August Strindberg - Bland franska bönder  
August Strindberg - Hemsöborna: Skärkarlsliv  
August Strindberg - Prosabitar från 1880-talet  
August Strindberg - Naturalistiska sorgespel  
August Strindberg - I havsbandet  
August Strindberg - Komdier och enaktare  
August Strindberg - En dåres försvarstal  
August Strindberg - Prosabitar från 1890-talet  
August Strindberg - Inferno  
August Strindberg - Legender  
August Strindberg - Till Damaskus  
August Strindberg - Vid högre rätt  
August Strindberg - Folkungasgan  
Hjalmar Bergman - Blå Blommor  
Hjalmar Bergman - Chefen fru Ingeborg  
Hjalmar Bergman - Clownen Jac  
Hjalmar Bergman - Dollar. Filmversion  
Hjalmar Bergman - Farmor och Vår Herre  
Hjalmar Bergman - Film  
Hjalmar Bergman - Hjalmar Bergman. Skrifter  
Hjalmar Bergman - Jag, Ljung och Medardus  
Hjalmar Bergman - Jonas och Helen. En studie över fantasi och ungdom  
Hjalmar Bergman - Knutmässa marknad  
Hjalmar Bergman - Komedier i Bergslagen. Två släkter  
Hjalmar Bergman - Labyrinten  
Hjalmar Bergman - Lotten Brenners ferier  
Hjalmar Bergman - Markurells i Wadköping  
Hjalmar Bergman - Solivro. Prins af Aeretanien  
Hjalmar Bergman - Swedenhielms. Komedi i fyra akter  
Carl Jonas Love Almqvist - Det går an. En tafla ur lifvet  
Carl Jonas Love Almqvist - Det går an  
Carl Jonas Love Almqvist - Dialog om sättet att sluta stycken  
Carl Jonas Love Almqvist - Drottningens juvelsmycke  
Carl Jonas Love Almqvist - Gudahataren ; Fragment  
Carl Jonas Love Almqvist -Herrarne på Ekolsund  
Carl Jonas Love Almqvist - Hvarför reser du?  
Carl Jonas Love Almqvist - Ladugårds-arrendet. En berättelse för folket  
Carl Jonas Love Almqvist - Svensk Rättstafnings-Lära  
Ada Rydström - Armfeltarne på Hunnerstad och deras närmaste  
Ada Rydström - Boken om Tjust  
Ada Rydström - Det gamla Adelsnäs

Ada Rydström - Stegeholm i Västervik  
Ada Rydström - Stegeholms slottsruin i ord och bild  
Emanuel Swedenborg - Apocalypsis revelata... Uppenbarelseboken avslöjad  
Emanuel Swedenborg - Den sanna kristna religionen  
Emanuel Swedenborg - Dædalius Hyperboreus  
Emanuel Swedenborg - Emanuelis Swedenborgii Opera Poetica  
Emanuel Swedenborg - Oförgripelige tanckar om svenska myntetz förnedring och förhöjning.  
Emanuel Swedenborg - Om Watnens högd Och Förra Werldens Starcka Ebb och Flod  
Emanuel Swedenborg - Swedenborgs drömmar 1744  
Emanuel Swedenborg - Äktenskapliga kärleken och dess motsats  
Gustaf af Geijerstam - Aldrig i Lifvet m. m.  
Gustaf af Geijerstam - Anteckningar om arbetarförhållanden i Stockholm  
Gustaf af Geijerstam - Boken om Lille-Bror  
Gustaf af Geijerstam - De gamla goda sagorna ånyo utgivna  
Gustaf af Geijerstam - Erik Grane. Uppsala-roman  
Gustaf af Geijerstam - Hur Lars Anders och Jan Anders byggde gärdesgård  
Gustaf af Geijerstam - Hur tankarna komma och gå. Intryck från Skansen  
Gustaf af Geijerstam - Hvad vill Lektor Personne? Ett genmäle  
Gustaf af Geijerstam - Karin Brandts dröm  
Gustaf af Geijerstam - Kronofogdens berättelser  
Gustaf af Geijerstam - Kulturkampen i Herjedalen. Från striden mellan lappar och bofaste  
Gustaf af Geijerstam - Lars Anders och Jan Anders och deras barn. En lantlig Comoedia i tre akter  
Gustaf af Geijerstam - Nils Tufvesson och hans moder  
Gustaf af Geijerstam - Nya brytningar : fem literära föredrag  
Gustaf af Geijerstam - Pastor Hallin  
Gustaf af Geijerstam - Samlade allmogebertättelser  
Gustaf af Geijerstam - Satirer och drömmar : dikter  
Gustaf af Geijerstam - Seklernas Nyårsnatt. Sagospel i en akt  
Gustaf af Geijerstam - Skogen och sjön : Berättelser  
Gustaf af Geijerstam - Ur samtiden. Literaturstudier  
Verner von Heidenstam - Karolinerna  
Verner von Heidenstam - Klassicitet och germanism  
Verner von Heidenstam - Stridsskrifter  
Verner von Heidenstam - Svenskarna och deras hövdingar  
Carl von Linné - Carl von Linnés ungdomsskrifter  
Carl von Linné - Caroli Linnæi Opera Hactenus Inedita. Flora Dalecarlica  
Carl von Linné - Methodus Avium Sveticarum  
Carl von Linné - Tal om märkvärdigheter uti insecterne  
Carl von Linné - Carl von Linnés anteckningar öfver Nemesis Divina  
HJ. Linderoth – Våra ortnamn och vad de lära oss  
Carl Snoilsky – Minnesteckningar och andra uppsatser  
Helena Nyblom – Det ringer! Skämt i en akt  
Helena Nyblom - Dollarpriessan en gammal violon  
Helena Nyblom – En Musik-kur Dramatisk scen för två personer  
Martin Koch – Mauritz - Självbiografisk Journalistik  
Bengt Lidforss – Kristendomen förr och nu  
Bengt Lidforss – August Strindberg och den uttärade nittiotalsreklamen  
Carl Pontus Wikner – Undersökningar angående den materialistiska världsåskådningen  
Carl Pontus Wikner - Undersökningar om Enhet och Mångfald  
Carl Pontus Wikner – Pontus Wikners Vittra Skrifter  
Carl Pontus Wikner – Filosofiens Historia  
Gustav Jansson – Segrare  
Erik Fahlman – Firman Åbergsonfirman Åbergson  
Elisa Fries – Anteckningar över Nemesis Divina  
Bonniers Litterära Magasin - Årgång I. 1932  
Bonniers Litterära Magasin - Årgång II. 1933  
Bonniers Litterära Magasin - Årgång V. 1936  
Bonniers Litterära Magasin - Årgång XI. 1942  
Bonniers Litterära Magasin - Årgång XII. 1943  
Bonniers Litterära Magasin - Årgång XIII. 1944



Bonniers Litterära Magasin - Årgång XIV. 1945  
Bonniers Litterära Magasin - Årgång XV. 1946  
Bonniers Litterära Magasin - Årgång XVI. 1947  
Bonniers Litterära Magasin - Årgång XVII. 1948  
Bonniers Litterära Magasin - Årgång XVIII. 1949

*Verk översatta till Svenska*

Charles Dickens - David Copperfield (översatt av H. Flygare)  
Charles Dickens - Den gamla antikvitetshandeln (översatt av C. J. Backman)  
Charles Dickens - En julsång på prosa  
Charles Dickens - Frelserens liv  
Charles Dickens - Jul-aftonen  
Charles Dickens - Liten Dorrit (översatt av La Strainiera)  
Charles Dickens - Moder och styfmoder  
Georg Lord Byron - Belägringen af Korinth  
Georg Lord Byron - Beppo. En venetiansk historia  
Georg Lord Byron - Bruden från Abydos  
Georg Lord Byron - Corsaren  
Georg Lord Byron - Giaurn, ett stycke af en turkisk berättelse  
Georg Lord Byron - Hebreiska Melodier  
Georg Lord Byron - Lara  
Georg Lord Byron - Mazeppa  
Georg Lord Byron - Ön eller Christian och hans stallbröder  
Jack London - Afgrundens folk  
Jack London - Afgrundsdjuret  
Jack London - All världens fiende med flere berättelser  
Jack London - Avgrundens folk (Norden)  
Jack London - Den lilla damen i det stora huset  
Jack London - Den röde guden  
Jack London - Efterskörd  
Jack London - En solens son  
Jack London - Fiskepatrullens berättelser  
Jack London - För mycket guld och andra berättelser från Klondyke  
Jack London - Före Adam  
Jack London - Glödande kol och andra noveller  
Jack London - Hans fäders Gud (Norden)  
Jack London - Hans fäders Gud och andra berättelser från Klondyke  
Jack London - Hawaji-noveller  
Jack London - Hedningen och andra berättelser från Söderhafsöarna  
Jack London - Jerry  
Jack London - John Finkelman  
Jack London - Järnhälen  
Jack London - Klasskamp  
Jack London - Kungen av Klondyke  
Jack London - Kvinnofejden. Pjäsen i tre akter  
Jack London - Kärlek till livet  
Jack London - Kärlekens väsen  
Jack London - Köldens barn (Norden)  
Jack London - Köldens barn  
Jack London - Martin Eden  
Jack London - Michael, Jerrys bror  
Jack London - Myteriet på "Helsingör"  
Jack London - Månansiktet och andra historier  
Jack London - Måndalen  
Jack London - Nattens barn och andra berättelser  
Jack London - Nigger Bellew  
Jack London - Nordlandets dotter. Berättelser  
Jack London - När gudarna skratta och andra historier  
Jack London - Ollonodlaren. Ett californiskt skogsdrama

Jack London - På kryss med "Blixten"  
Jack London - På luffarstråt (Norden)  
Jack London - På långfärd med Snark  
Jack London - På tro och loven (Norden)  
Jack London - Revolution och andra essayer  
Jack London - Röda pesten  
Jack London - Skriet från vildmarken  
Jack London - Solguld och andra berättelser  
Jack London - Spelet  
Jack London - Strejkbrytaren  
Jack London - Stöld. Pjäs i fyra akter  
Jack London - Söderhavsberättelser  
Jack London - Tre hjärtan  
Jack London - Tvångströjan  
Jack London - Vagabondliv. Ungdomsminnen  
Jack London - Vargens son. Skildringar från polartrakterna  
Jack London - Varghunden  
Jack London - Äventyr  
Fjodor Dostojevskij - Arma människor (översatt av Alfred Jensen)  
Fjodor Dostojevskij - Vita Nätter (översatt av Alfred Jensen)  
Fjodor Dostojevskij - Vårdinnan (översatt av Alfred Jensen)  
Fjodor Dostojevskij - En Skriftställares dagbok (översatt av Alfred Jensen)  
Fjodor Dostojevskij - Krotkaja (översatt av Gerhard Gran)  
Leo Tolstoj - Berättelser för folket. (översatt av E. Weer)  
Leo Tolstoj - Till krig emot kriget  
Leo Tolstoj - Furst Nechljudo (översatt av Gunnar Skans)  
Leo Tolstoj - Efterlämnade skrifter III (översatt av Walborg Hedberg)  
Leo Tolstoj - Ivan Iljitj död (översatt av E. Weer)  
Leo Tolstoj - Kosackerna (översatt av E. Weer)  
Leo Tolstoj - Mörkrets Makt (översatt av Walborg Hedberg)  
Leo Tolstoj - Döda icke (översatt av Ellen Rydelius)  
Leo Tolstoj - Anna Karenina (översatt av Oscar Nachman)  
Leo Tolstoj - Kreuzersonaten (översatt av Gunnar Skans)  
Nikolaj Gogol - Berättelser (översatt av AF K. E. Peterson)  
Nikolaj Gogol - Revisorn (översatt av Erik Nordenström)  
Nikolaj Gogol - Taras Bulba (översatt av Jarl Hemmer)  
Victor Hugo - Ringaren i Notre-Dame (översatt av Osvar Nachman)  
Victor Hugo - Hafvets Arbetare  
Johann Wolfgang von Goethe – En skön Själs Bekännelser (översatt av Valborg Hedberg)  
Johann Wolfgang von Goethe – Valfrändskap  
Jules Verne - Den hemlighetsfulla ön (översatt av Gustaf Nordenskiöld)  
Jules Verne - En öfvervinting i polar-isen  
Jules Verne - Från jorden till månen  
Jules Verne - Ingenjör Roburs luftfärd  
Jules Verne - Jorden runt på 80 dagar  
Jules Verne - Till Jordens medelpunkt  
Jules Verne - Tsarens kurir: från Moskva till Irkutsk  
Jules Verne - Åtta hundra mil på Amasonfloden

## Datamängd 3

*En submängd ur Datamängd 2 – enbart lyrik:*

Dan Andersson - Dan Anderssons samlade skrifter: minnesupplaga  
Dan Andersson – Dikter i Det kallas vidskepelse samt Brooklandsvägen  
Dan Andersson – Dikter  
Dan Andersson – Efterlämnade dikter  
Dan Andersson – Kolvaktarens visor  
Dan Andersson – Svarta Ballader  
Dan Andersson – Visor och ballader  
Edith Södergran – Brokiga iakttagelser  
Edith Södergran – Dikter  
Edith Södergran – Framtidens skugga  
Edith Södergran – Landet som icke är  
Edith Södergran – Rosenaltaret  
Edith Södergran – Septemberlyan  
Edith Södergran - tankar om naturen  
Gustaf Fröding – Fyra dikter  
Gustaf Fröding – Gitarr och dragharmonika  
Gustaf Fröding – Gralstänk  
Gustaf Fröding – Grillfångeri  
Gustaf Fröding – Mattoidens sånger  
Gustaf Fröding – Nya dikter  
Gustaf Fröding – Om Lifsmonader  
Gustaf Fröding – Skrifter, jubiliemsupplaga  
Gustaf Fröding – Stänk och flikar  
Erik Axel Karlfeldt – Efterskörd  
Erik Axel Karlfeldt – Erik Axel Karlfeldts dikter  
Erik Axel Karlfeldt – Flora och Bellona  
Erik Axel Karlfeldt – Flora och Pomona  
Erik Axel Karlfeldt - Fridolins lustgård och Dalmålningar på rim  
Erik Axel Karlfeldt - Fridolins lustgård och andra dikter  
Erik Axel Karlfeldt – Hösttorn  
Erik Axel Karlfeldt – Vildmarks och kärleksvisor  
Karin Boye – De sju dödssynderna och andra efterlämnade dikter  
Karin Boye – För trädets skull  
Karin Boye – Gömda land  
Karin Boye – Härdarna  
Karin Boye – Moln  
Erik Johan Stagnelius – Lyriska dikter intill tiden omkring 1818  
Erik Johan Stagnelius – Strödda dikter efter tiden omkring 1818  
August Strindberg - Dikter på vers och prosa.

## Datamängd 4

*En submängd ur Datamängd 2 – enbart Karin Boyes lyrik:*

Karin Boye – De sju dödssynderna och andra efterlämnade dikter  
Karin Boye – För trädets skull  
Karin Boye – Gömda land  
Karin Boye – Härdarna  
Karin Boye – Moln