

Namespace Demonstration Scenario

Karina van Dalen-Oskam
16 december 2013



1. GENERAL INTRODUCTION

This demonstration scenario is meant to be a guide for the demonstrator applications that resulted from the CLARIN Namespace project. The demonstrator applications are: a search interface to the curated corpus data, a barcode browser, the Namespace visualizer, and the Namespace NER tagger. It will not go into the Namespace Sandbox, which is for authenticated users only.

The website www.namespace.nl can be used for several tasks related to named entity recognition, named entity resolution and the visualization of results from earlier analyses of works of fiction originally in Dutch or translated into Dutch published during the last decades. Several text collections have been enriched, but not all of these are freely accessible through the website. More information about the corpora can be found on http://blog.namespace.nl/?page_id=85. During the processing of the corpora we found they included many texts that are not fiction but non-fiction. We decided to leave them in the corpus, since it would take more time to remove them than to just process them as well.

One general remark: the web services have to deal with quite a lot of data and are therefore not very fast. Be patient.

2. SEARCH AND VISUALIZE DATA

To **search** the proper names in already tagged texts, go to <http://search.namespace.nl/>. You can address your search in one or all of the corpora, but only PG (project Gutenberg) and SoNaR will also link to the complete text. To walk you through one possible route: Select SoNaR, and “alleen nette alinea’s” (“only well-shaped paragraphs”). Suppose you want to know how many novels have a character in it called “Jan”. Fill in “Jan” in the field “persoonsnaam” (personal name) and click on “zoek” (search). The search yields 23 hits that are easily inspected below the search interface, through the keyword in context presentation. Most of the hits use “Jan” not as the name of a character.

The acronyms at the far right after each hit refer to **GB**: Google Books, **LT**: LibraryThing, **WC**: WorldCat, and **OL**: Open Library. The green link below the title leads to the XML output.

1. 2009 *Zeeuw, Ron de: De zwarte lady MacBeth* 9789058314963 [GB I](#)
... Jan overnachtte bij ...
(C) genre: — [nl.ns.d.9789058314963.c](#)
2. 2010 *Zeeuw, Ron de: Schaduwgrens* 9789058315281 [GB LT WC OL](#)
... was en hoopte Jan aan zijn kan ...
(C) genre: — [nl.ns.d.9789058315281.c](#)
3. 2004 *Robbert-Jan Henkes en Erik Bindervoet: Bloemsdag*
... Kale Bertus , Jan met de Handj ...
(C) genre: — [nl.ns.d.f0b1145e35739d4b526bfce1c2993f21508b991d.c](#)
4. *Ivo de Wijs (bewerking van origineel van Hildebrand (pseudoniem van 'Jemig , Jan I dat is ...*
(C) genre: — [nl.ns.d.f53a91f26ddaa7d598b2d6be9c74c47c5bf7d4e5.c](#)
5. *Adamoli, Vida: Lisa's huishouden* 9058313492 [GB LT WC OL](#)
... De Lange Jan met zijn opg ...
(C) genre: — [nl.ns.d.9058313492.c](#)
6. *Molist, Jorge: De ring* 9058313786 [GB LT WC OL](#)
... avond van Sint- Jan , die van ...
(C) genre: — [nl.ns.d.9058313786.c](#)
7. 1986 *Carlo Fruttero and Franco Lucentini: De minnaar zonder verbl*
... Jan en alleman , ...
(C) genre: — [nl.ns.d.8bac15582888c77d26f2185d32fc77fb955e339d.c](#)

The first hit has a character named "Jan". To further inspect this novel, click on the hyperlinked title *De zwarte lady MacBeth*. This shows the full text to the right, with an overview of the most frequent names in the left hand column:

<http://resolver.namescape.nl/nl.ns.d.9789058314963.c?view=html&q=>



Bibliografische gegevens

Titel De zwarte lady MacBeth

Auteur Zeeuw, Ron de

ISBN 9789058314963

Jaar 2009

Synopsis

persoon (1952)

JAN HOUTER (364)

JAN HOUTER SJEF (227)

CARLO KOOLMEIJER (176)

plaats (630)

NEDERLAND (86)

SURINAME (62)

PARAMARIBO (47)

organisatie (117)

AIR SURINAM (12)

DE MR LIM A POSTRAAT (9)

VREEMDELINGENPOLITIE (7)

overig (9)

JAN EVERTSENSTRAAT (2)

LUCAS ANDREASZIEKENHUIS (2)

HAROLD MAHANGOE (1)

De zwarte lady **MacBeth**

Aan Christina Maria Eype

Sex is too wonderful to share it with anybody else

- Philip Larkin

1

Zoals altijd begon hij het proces-verbaal met :

' Ik , **Jan Houter** , hoofdinspecteur van de regiopolitie , dienstdoende bij de centrale recherche van de regiopolitie **Amsterdam-Amstelland** , verklaar het volgende ... '

Wat volgde waren een paar eenvoudig in te vullen onderdelen zoals personalia , melding en tijdstip van overlijden .

Daarna dacht **Houter** aan de komende marathon van **Amsterdam** , aan zijn grootmoeder , aan zijn broer **Jan-Jaap** , aan zijn tweeënveertigduizend gulden spaargeld , aan zijn huis , dat sinds de aankoop in 1990 in zes jaar in waarde was verdubbeld , aan zijn ex-vrouw , en aan het boek *The long goodbye* van **Raymond Chandler** , dat hij voor de derde maal herlas en waarin hij ontzettend zin had door te gaan . Hij schudde het hoofd en sperde de ogen wijd open , wat soms hielp om weer bij het werk terug te komen , en typte :

' Kort na het overlijden werd , in het bijzijn van collega **P.M.J.** van Vlodrop , het stoffelijk overschot , **Carlo Petrus Koolmeijer** bovengenoemd , geïdentificeerd door zijn echtgenote , genaamd **Sarita Koelfat** , geboren te **Nieuw-Nickerie** , **Suriname** , op 24 januari 1962 , wonende **Tefelenstraat 214** , te 1102 GA **Amsterdam Zuid-Oost** . '

Zoals altijd werd hij toen kwaad . Zijn bureau lag vol aantekeningen , kaffeelietjes

On the top left, the bibliographic reference is given. Below, the amount of personal names (1952), place names (630) and organizations (117) is mentioned, each category with the three most frequent name in that category.

The **Barcode Browser** gives another entrance to the names in the different corpora. Go to <http://barcode-browser.namescape.nl/index.xql> and search in corpus Sanders for the personal name "Jan". After a couple of minutes you'll see a large amount of columns, one for every book which contains "Jan". The length of the columns visualizes the length of the books. Each red band marks an occurrence of "Jan". Clicking on such a red band shows the paragraph in which "Jan" occurs, with all other names highlighted as well (including the colour coding for each name category). Not all novels fit the screen; there is a scroll bar below the columns. Mouse-over the blue header of a column shows author and title of the book visualized in that column. The Jan-example can be found here: <http://barcode-browser.namescape.nl/index.xql?q=Jan&author=&title=&startdate=&enddate=&collection=sanders>

The **Namescape Visualizer** helps to give an overview of the names found in a text and shows the co-occurrence of names in paragraphs. Suppose you would like to know the onymic landscape (the "landscape" of proper names) in the novel *De vergaderzaal* by A. Alberts. When you are in <http://visualizer.namescape.nl/> then choose "Bekijk alle boeken" ("browse all books"), and choose "Auteur" ("author"). Click on the A. Alberts' novel *De vergaderzaal*. The tool then shows the twenty most frequent names (as recognized by an earlier tagger), with an automatically generated link to Wikipedia (<http://visualizer.namescape.nl/book/508>). Three different visualizations of the data are presented.

Bundle

<http://visualizer.namescape.nl/graph/index/type/bundle/bookid/508> shows the characters (personal names) and the level of co-occurrence in paragraphs in the complete novel. In the bar at the top, you can opt for including more than twenty names by sliding the bar to the right. You can also adjust the “tension” in the bundle. *De vergaderzaal* is a rather short novel, which does not present very exciting pictures.

Matrix

<http://visualizer.namescape.nl/graph/index/type/matrix/bookid/508> presents a matrix graph. The amount of names included can be adjusted. You can also readjust the matrix based on frequency and on cluster-level, which is done dynamically. This shows better which names co-occur most and in which clusters of characters.

Barcode

<http://visualizer.namescape.nl/graph/index/type/barcode/bookid/508> gives a barcode graph, indicating where the different names (twenty or more, as in the other visualizations) occur in the text, from the first to the last page (from left to right). This gives another perspective on how the role of characters for instance changes during the story.

3. TAG YOUR OWN TEXT

Namescape has two tools for named entity recognition that you can apply to your own digital texts. You can submit the text in several formats: text, html, epub, Word, TEI. Go to

<http://ner.namescape.nl/namescape/tagger> for the very user-friendly interface. Choose the file of your choice from your own computer or from the web, select the relevant input format (e.g. txt) and output. Choose the NER tagger you want to use. The output “Styled” shows the text on screen, with the names highlighted. At the top, a set of links leads to an overview of all the names in different categories. Clicking on a name leads to the paragraph in which the name occurs, again with the names highlighted. The output “Link” gives a link to an XML-TEI representation of the names and their occurrences. The output “Raw” presents the XML-TEI output on screen. An example of the output “Styled” goes below.

unknown

idno: SIDNO
author: unknown
title:
yearFrom: unknown
yearTo: unknown
source: unknown

Show/hide name information

Persons, Locations, Organisations, Misc, Cooccurrence graph

Ge-eerde heer, zeer ge-achte voogd!

Dat was een blyde Postdag voor my. Een Brief van myn ge-eerden Voogd. Waarlyk, ik heb geschreid, ziende hoe veel belang gy in my naam: doch dat zes maanden uit blyven! daar lag al myn vreugd in 't voetzant. Wel, myn allerleefste Voogd, ik kan het hier geen zes weken langer uithouden; zo als ik ook evenwel behandelde woord, maar ik kan 't niet half schryven; zo gy, myn Heer, hier waart, gy zoudt het my toestaan. Och, zo waar, ik heb u geen **♦♦♦** in jokkentie, hoe klein ook, op den zeeuw gespelt. Foei, myn Heer, zou ik liegen? dan was ik een zeer slecht meise, en verliende dat gy my bekeeft. Ik ben niet alleen de slavin van Tantes grillen, maar ik word ook geringelooft door een oude jelyke zotte meid, die om Tante te behagen, my del maakt. De Juffrouw, daar ik gaarne by zoude invomen, is de ongelukkige weduw van een fatsoenlyk man, die niet dan ordentelyke Dames tegert. Eene myner oude schooldaartjes is daar reeds erigen tyd geweest, en pryst de Juffrouw heel zeer. Daar zyn nog twee andere Dames ook **Vrij**, vrolyk en onbesnoet te leven, dit is myn enig cogmerk; en is dat beispelyk? By Tante kan ik niet blyven, zo ik my niet tot huichdary wil verlagen, eenne ondiepl, die allerasschuwlykst voor my is; en waar aan ik my zeker noot zal te buitengaan. Ik beveel my in uwe gurst. Ik zal my in allen opzichte altoos zo pogen te gedragen, dat gy voldaan zyt, maar by Tante kan ik niet blyven. Laat my toe, dit nogmaal te zeggen. Wat ben ik blyde met de my toegezegde Muziek! Ik zing al in voorraad, of Wat zal die fraai zyn; **Mozglyk** is er wel van **Roussens's Compositie** by' duizendmaal dank. Ik hoop al die stukken u eens, onder het rooken van een Pypje, voor te spelen. Maar, denk eens aan, myn Heer **Blaukaart**, daar wil Tante niet hebben dat ik speel, dan naar eenre koflige zuszery♦♦♦; en lieve Heer, ik speel evenwel zo graag, en ik heb zulke mooie Carata's. Mag ik u bidden, myn Heer, zendt het Falje aan Tantes huis niet; het ging wis en waarlyk op 't vuur; ik zal hier een adresje insluiten. Ik bidde den goeden **Hemel** alle daag voor u, en dat ik u gezond en vrolyk moge weder zien, my zelf gelukkig rekenende van te zyn.

Uwe liefhebbende **Puppl** en **Dienares**,
Sara Burgerhart

Auteur: Sara Burgerhart
Moochvanden...

If in the horizontal bar right above the text you click on “persons”, you’ll see the list of named entities that the tool recognized as being persons; clicking on “persons” again closes the list so you can then select another option. The list contains all different names of a person grouped under the

entity. This gives an easy overview of the characters mentioned in the text and even of their relative importance through the over-all frequencies of the proper names and the entities.

yearTo: unknown
source: unknown

Show/hide name information

[Persons](#), [Locations](#), [Organisations](#), [Misc](#), [Cooccurrence graph](#)

[LETJE BRUNIER](#) (97)
Letjes (LETJE) (1)
Brunier (BRUNIER) (12)
Letje Brunier (LETJE BRUNIER) (1)
Letje (LETJE) (83)

[BUIGZAAM](#) (77)
Buigzaam's (BUIGZAAM) (1)
Buigzaam (BUIGZAAM) (76)

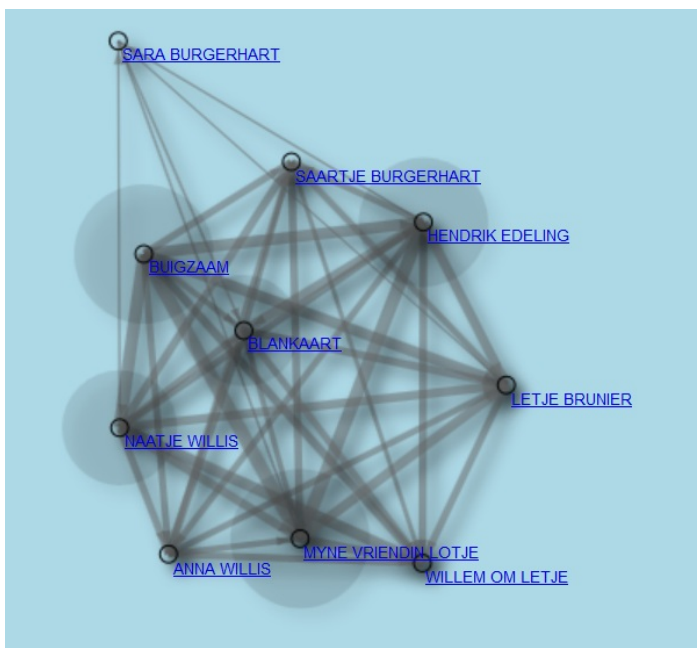
[MYNE VRIENDIN LOTJE](#) (60)
Myne Vriendin Lotje (MYNE VRIENDIN LOTJE) (1)
Lotjes (LOTJE) (1)
Lotje (LOTJE) (58)

[JUFFROUW HARTOG](#) (56)
Juffrouw Hartog (JUFFROUW HARTOG) (54)
Juffrouw (JUFFROUW) (2)

[NAATJE WILLIS](#) (50)
Naatjes (NAATJE) (1)
Naatje (NAATJE) (48)

javascript:searchNym("nym.person.81",10,false) [AATJE WILLIS] (1)

The right-hand option, co-occurrence graph, shows a co-occurrence graph of all the persons in the text, to explore the relations between them:



More about the tagger can be found in the help:
<http://ner.namescape.nl/namescape/web/help.html>