1. GENERAL INTRODUCTION

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

The website www.namescape.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.namescape.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.namescape.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.
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:

```
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=
```
The **Barcode Browser** gives another entrance to the names in the different corpora. Go to [http://barcode-browser.namescape.nl/index.xql](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](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/](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](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.

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.

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.nameescape.nl/nameescape/web/help.html