

Updated: November 7, 2005

NOME USER GUIDE

Contents

Introduction	2
What is NOME?	2
The technology	2
Use and distribution	3
Acknowledgements	3
Installation and update	4
Requirements	4
Download	4
Installation	4
Uninstall	5
NOME's basic detection logic	6
Anonymizing with NOME	7
Launching NOME	7
Using the replacement list	8
Browsing the document	9
Adding or removing terms within a session	10
Changing the suggested replacements	11
Replacing names	13
Saving the replacements	15
Refining NOME's basic detection logic	17
Inclusion list	17
Titles list	17
Exclusion list	17
Changing and creating lists	18
Other settings	20
Recommended workflow	21

Introduction

What is NOME?

NOME is a software application that assists editors in anonymizing the names of certain persons in a document. It helps editors select which persons should be anonymized, locate occurrences of proper names and replace them with initials or other sets of characters. NOME's contribution to the anonymization process, which should remain driven by a trained individual, is to help achieve greater accuracy and efficiency. The assistance provided by NOME can be summarized as follows:

- NOME detects proper names in a document.
- NOME allows the list of detected proper names to be consulted and reviewed.
- Detected terms can be browsed in the document, revealing their specific context, thus helping the editor to determine whether the occurrences refer to a single person.
- NOME suggests replacement initials.
- NOME offers the possibility of systematically replacing all occurrences of a specific term.
- NOME allows replacements to be reviewed and then accepted or rejected.

At every step, NOME's suggestions can be rejected or modified by the editor, who ultimately controls the editing process.

The technology

NOME is a "scripted program" embedded in MS Word. When activated, it displays a specially designed panel in which potential replacements can be displayed, chosen, modified and executed. Potential replacements are found automatically and executed systematically throughout the whole document. They can be approved by the user, and even reversed, using the MS Word "Track changes" tool. Usual Word commands remain available to the user, so NOME can be integrated easily into any office automation environment.

Use and distribution

NOME is the result of collaborative effort by the University of Montreal's LexUM (Faculty of Law) and RALI (Department of Computer Science). It was designed to simplify some technical aspects of the process of anonymizing names in published judgments, in compliance with legal bans on publication. The tool was first adopted by the editorial personnel of the Canadian Legal Information Institute's website (www.canlii.org). However, the developers' main objective is to make the software available to any court, tribunal, bar association or similar decisions issuing institutions, which are in the best position to perform this type of editorial work since they have better knowledge of applicable bans on publication. In order to make it available as widely as possible, NOME is distributed under an open source license. The license entitles the recipients to freely use, distribute and modify the software to better meet their specific needs. Improvements to NOME can be shared by sending an e-mail to CanLII's editorial team at edit@canlii.org.

Acknowledgements

The core funding for the development of NOME was provided by the Canadian Legal Information Institute (CanLII). CanLII is a not-for-profit corporation funded by all Canadian legal professionals through the Federation of Law Societies of Canada. CanLII's mandate is to improve accessibility to Canadian legal materials for the benefit of the legal profession and the public as well. A previous version of NOME was developed with the support of the Société québécoise d'information juridique (SOQUIJ).

The collaborative efforts of LexUM and RALI were supported by an infrastructure funding grant from the Canadian Foundation for Innovation, through the TAPoR consortium.

NOME's functional requirements were established by LexUM's editorial team. The software application was developed by RALI's Ngoc Tran Nguyen and Luc Plamondon, under the supervision of Guy Lapalme.

Installation and update

Requirements

To install NOME on your computer, you need:

- Microsoft Windows XP as your operating system;
- Microsoft Word 2002 (included in Office 10) or higher; and
- Administrator's rights for your workstation.

Note: If you are using Nome on your own personal computer, you normally have administrator's rights. In an office network, however, you may need the intervention of the network's system administrator to install NOME on your workstation.

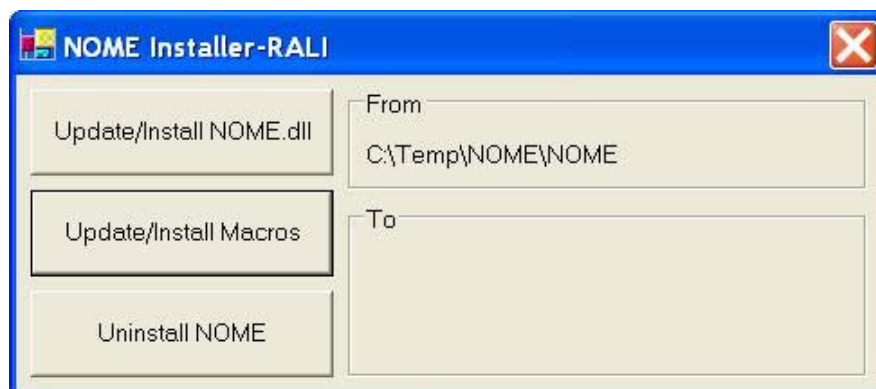
Download

The latest version of NOME is available from RALI at <http://rali.iro.umontreal.ca/Nome>. Download the ZIP file and save it on your workstation. Double-click on the ZIP file and run the extract command. The installation files will be extracted under a new subfolder named NOME.

Installation

Before installing Nome, make sure to quit Microsoft Word, and, preferably, all running software.

- The NOME subfolder contains a file called "installer.exe". Double-click on it and the following dialog box appears:



For the next step, administrator rights are needed.

- NOME.dll must be installed by clicking on Update/Install NOME.dll.
- NOME Macros can be installed by any user by clicking on Update/Install Macros.

Uninstall

NOME can be uninstalled by either a user or an administrator by clicking on Uninstall NOME.

- Launch the program by double-clicking on "installer.exe";
- Then, click on "uninstall NOME".

NOME's basic detection logic

The present section focuses on NOME's basic detection logic, with which the user should be aware of when working with NOME.

The basic logic of NOME is that every sequence of two or more capitalized words that are not separated by a dot is considered a proper name.

In a document, ***Beverly Smith*** would be identified as a proper name. According to the same logic, ***Registered Retirement Income Fund*** and ***Canadian Imperial Bank*** would also be considered proper names.

Every component of a term identified as a proper name is further interpreted as an individual proper name. As a result, for example, if a victim in a case is randomly referred to by either her family name or her first name, all such occurrences will be detected by NOME, as long as the person is referred to by her full name at least once.

Since the sequence ***Beverly Smith*** is identified as a proper name, every term contained therein – ***Beverly*** and ***Smith*** - is also interpreted as an individual, independent proper name.


This basic logic does not rely on a specific dictionary of proper names. It can sometimes produce results that will need to be ignored and omit names that should be detected.

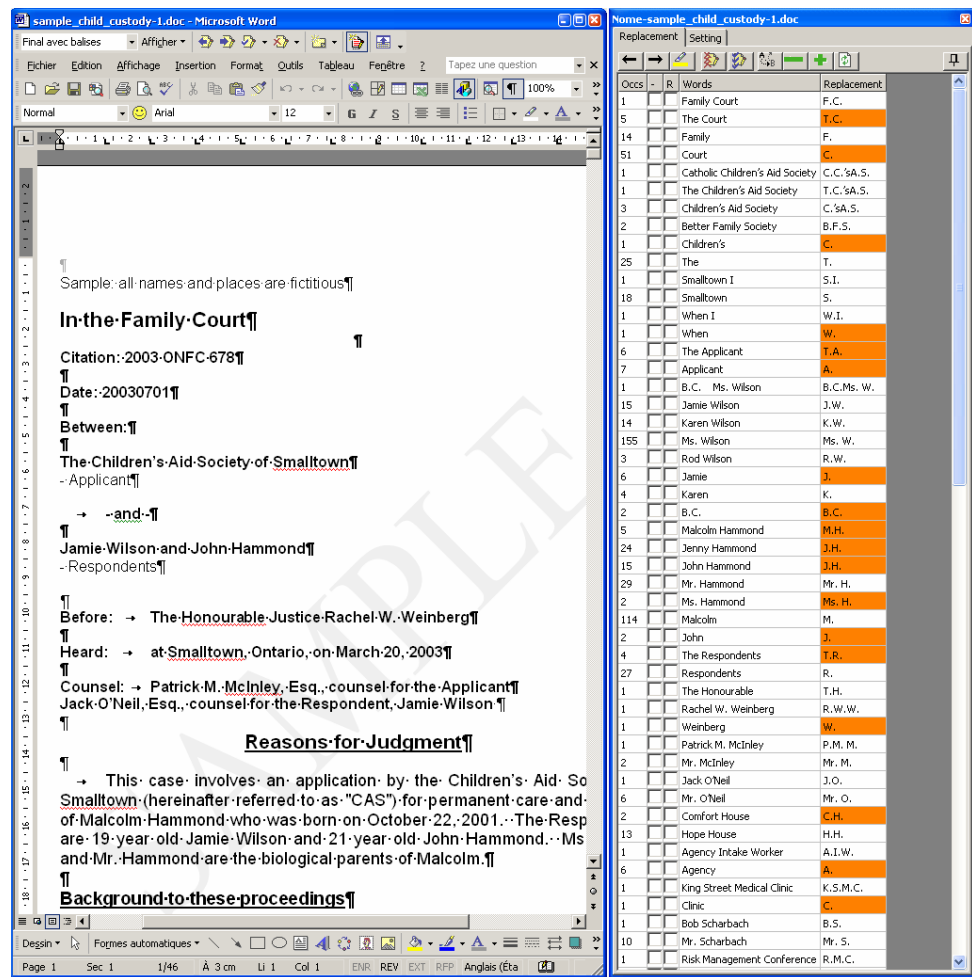
The sequences ***When I***, ***In October*** and ***Provincial Court*** would be considered proper names of persons because they fulfill the basic requirement of containing two or more terms starting with capital letters.

This manual will later describe how this basic logic can be refined by the customization of permanent exclusion and inclusion lists of terms or changed for a single document by manual addition or removal of occurrences within a specific session.

Anonymizing with NOME

Launching NOME

A NOME session is launched by opening the target document and then clicking on the NOME button (), which appears on the MS Word task bar. Upon launch, the content of the document is analyzed. Then the NOME interface window appears on the right hand side of the document. The interface window already contains the result of the document analysis displaying a replacement list, that is, occurrences of terms identified as proper names with their corresponding suggested replacements. (Fig. 1)

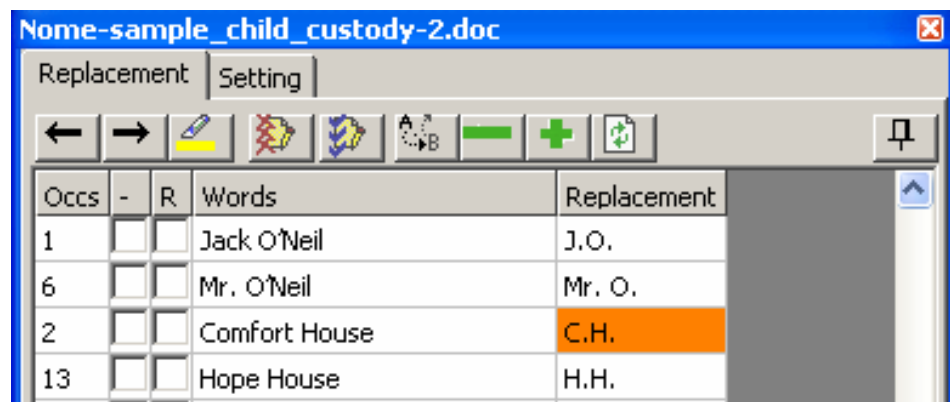


[Figure 1: The NOME interface window]

Using the replacement list

Upon completion of the automated document analysis and the ensuing detection of proper names, NOME creates a replacement list that can be edited by the user.

Every line from the replacement list contains the *number of occurrences*, the *proper names* themselves and *suggested replacement initials*. (Fig. 2) The list can be sorted in accordance with any of these elements by clicking on the cell at the top of the relevant column.



Occs	-	R	Words	Replacement
1	<input type="checkbox"/>	<input type="checkbox"/>	Jack O'Neil	J.O.
6	<input type="checkbox"/>	<input type="checkbox"/>	Mr. O'Neil	Mr. O.
2	<input type="checkbox"/>	<input type="checkbox"/>	Comfort House	C.H.
13	<input type="checkbox"/>	<input type="checkbox"/>	Hope House	H.H.


[Figure 2: The replacement list]

Since the detected proper names are broken down into their components, individual entries are created for both full names and components.



The list can contain *John Smith*, *John* and *Smith* as three separate entries.

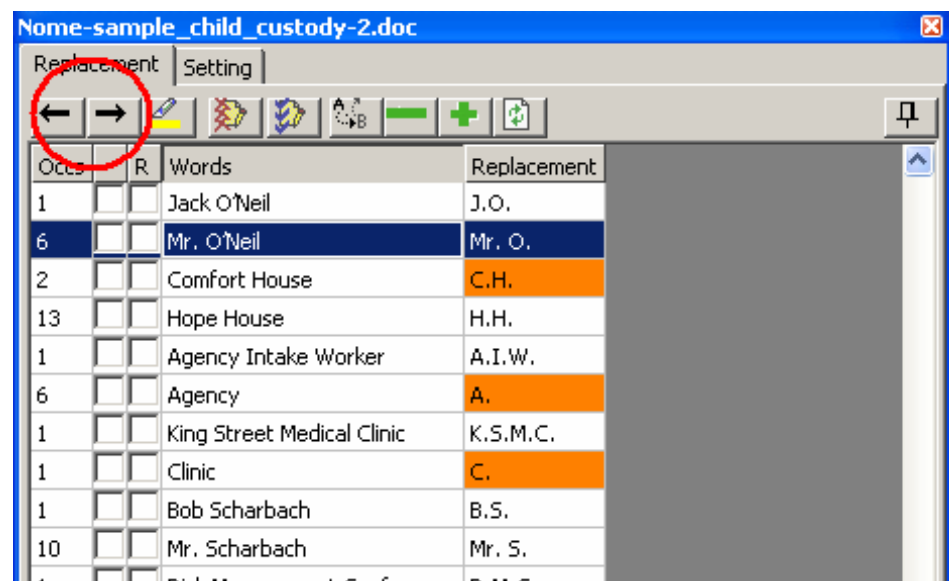
Every individual entry has to be treated separately. Entries cannot be merged. When two or more entries relate to the same person, they should receive identical but separate treatment. Entries relating to the same person can be treated identically by either ignoring them all or replacing them all with the same initials.

When the following three entries - **Jane Smith**, **Jane** and **Mrs. Smith** - correspond to the same person, replacing or ignoring **Jane** will have an impact on the occurrences of **Jane** alone and NOT on the occurrences of **Jane** in **Jane Smith**.

The replacement list can be refreshed by clicking on . NOME will then reanalyze the current version of the document and provide a new replacement list. Please note that this new replacement list will still contain terms that were manually added but will also include terms that were manually removed from the previous list.

Browsing the document

In order to correctly assess what names should be replaced or what names correspond to the same person in the document, the user may want to check them in their context. The replacement list allows the user to easily do so by browsing through the parts of the document in which a specific occurrence of a word appears. In order to select the corresponding line, the user simply clicks on the cell containing the term. Then the user can navigate through this term's occurrences in the document text by clicking on  or . (Fig. 3)




[Figure 3: Navigating through occurrences]

Adding or removing terms within a session


The replacement list allows for the manual addition or removal of terms within a session, which is often unavoidable. These modifications apply only to the specific document being edited within a session. For adjustments using inclusion and exclusion lists that will apply to whole sets of documents, please refer to the next section, “Refining NOME’s basic detection logic”, below.


Adding a term

When a proper name is not recognized by NOME, it can be added for a specific session by selecting it in the document text and then clicking on . The term added will appear in a different color at the end of the replacement list.

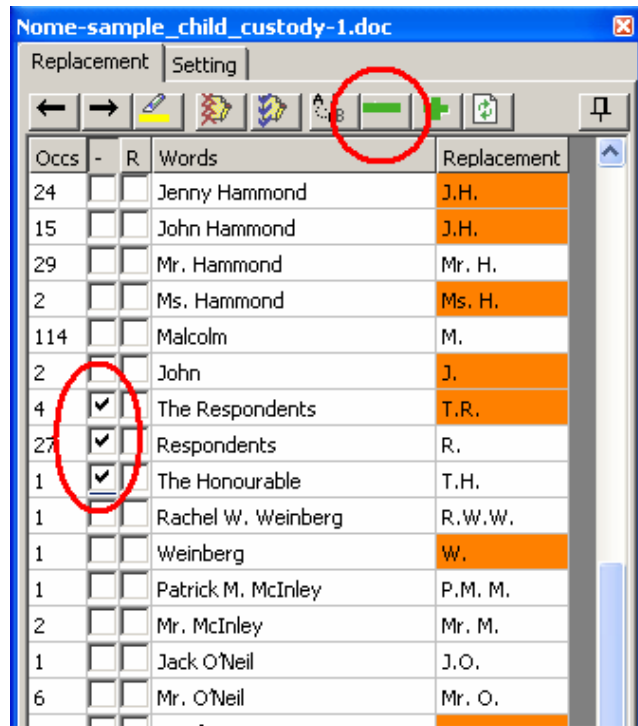
If a person is referred to solely as **Jane**, NOME will ignore the occurrences of **Jane** in the document. This may require manual addition of the term so that NOME can consider it a proper name.

Ignoring a term

NOME may detect terms that are irrelevant. Such a term can be removed by checking the corresponding box in the *minus* (-) column and then clicking on .

As a first step in a session, it is recommended that the list of names in the replacement list be shortened by removing those that seem clearly irrelevant, in order to obtain a better view of the proper names that should be anonymized. In such a scenario, the user can scroll down the list and check the corresponding boxes, and then click on  to remove all the selected terms at once. (Fig. 4)

Generally, terms such as ***In Morgentaler, On April 24, The Applicant***, etc. can be ignored from the start. Other terms, such as judges’ and counsel names, can also be removed at a later stage in the document analysis.



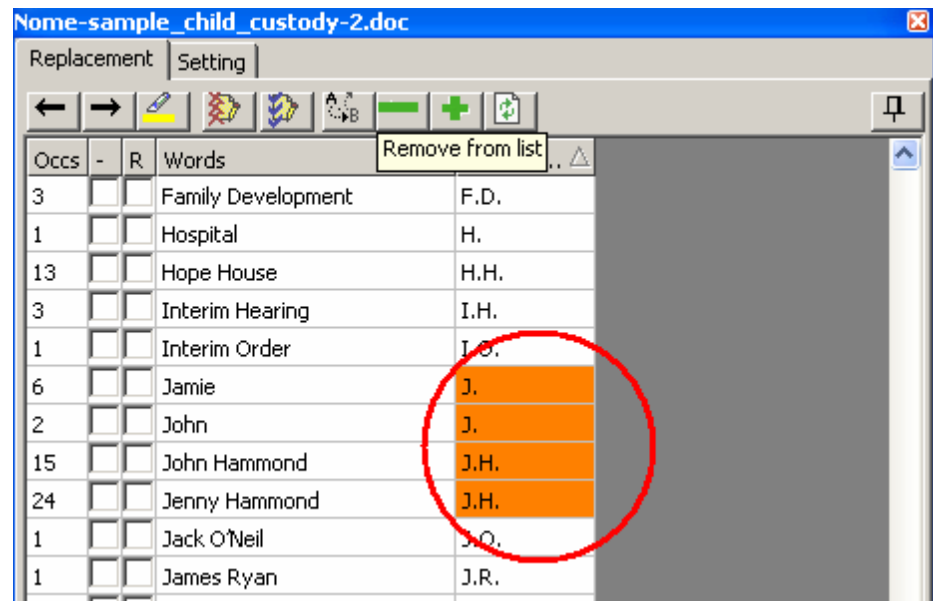
[Figure 4: Ignoring terms in a session]

Changing the suggested replacements

The suggested replacements can be changed by clicking on the corresponding cell and by typing the proper replacement characters in the Replacement column.

The user may conclude that two or more individual entries relate to the same person, e.g. **Jane**, **Jane Smith** and **Mrs. Smith**. In such a case, the user may decide to use the same initials – **J.S.** - for these separate entries. Thus, the initials set by default by NOME – **J.**, **J.S.** and **Mrs. S.** can be replaced manually by **J.S.**

When the same replacement initials are used in different list entries, NOME will highlight the entries in order to alert the user of a potential conflict of initials. (Fig. 5) This feature does not prevent the user from making a replacement and can be ignored if not relevant.



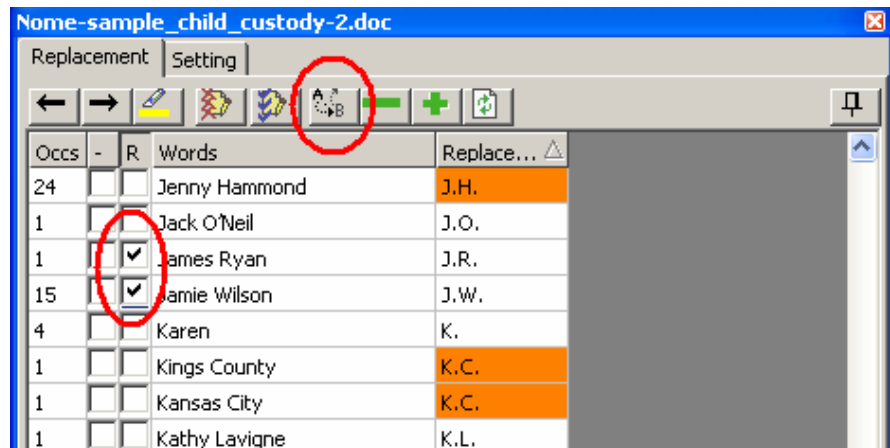
[Figure 5: Potential conflict of initials]

The same initials should not be used for replacing the names of different persons, as this could seriously affect the readability of the document.


Replacing names

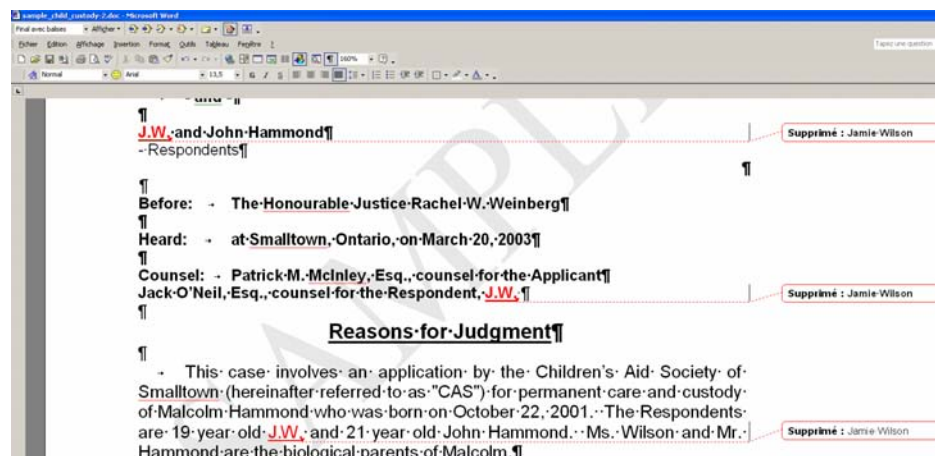
Making a replacement

In order to make a replacement, the user first checks the replacement column box belonging to the entry of the term to be replaced. (Fig. 6)





[Figure 6: Making a replacement]

The user can later execute the replacement by clicking on . NOME will then replace all occurrences of the selected term by the corresponding initials. NOME activates the MS Word Track Changes function, displaying changes as requiring approval to become final. (Fig. 7)



[Figure 7: The Tracked Changes]

Consulting replacements

In order to consult a replaced term in the document, the user has to select the corresponding line in the replacement list and then browse through the replacements of this term in the document by clicking on  or . (Fig. 5)

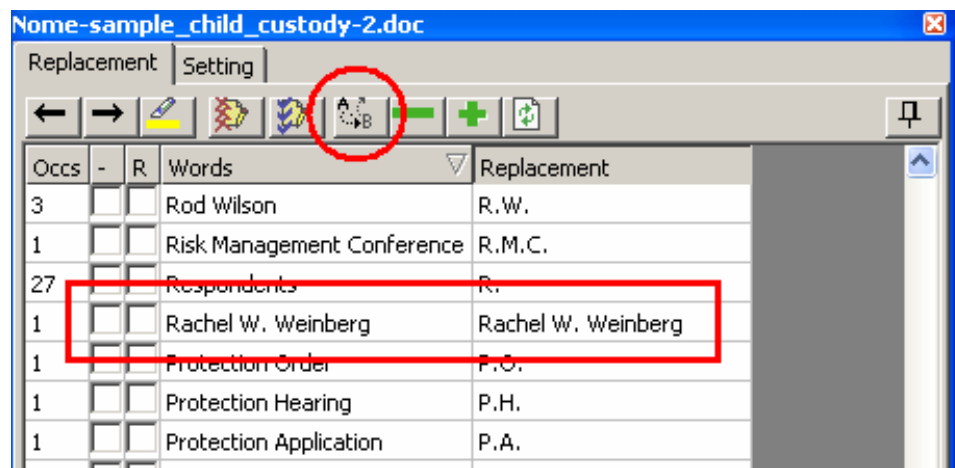
Changing or undoing a replacement

All replacements of a term can be changed by performing another replacement of the term in question.

If **Jane Smith** was replaced by **J.S.**, the initials can be changed to **J.Sm.** by modifying the initials in the line corresponding to the proper name Jane Smith and then re-executing the replacement.

This method can also be used to undo a replacement. (Fig. 8)

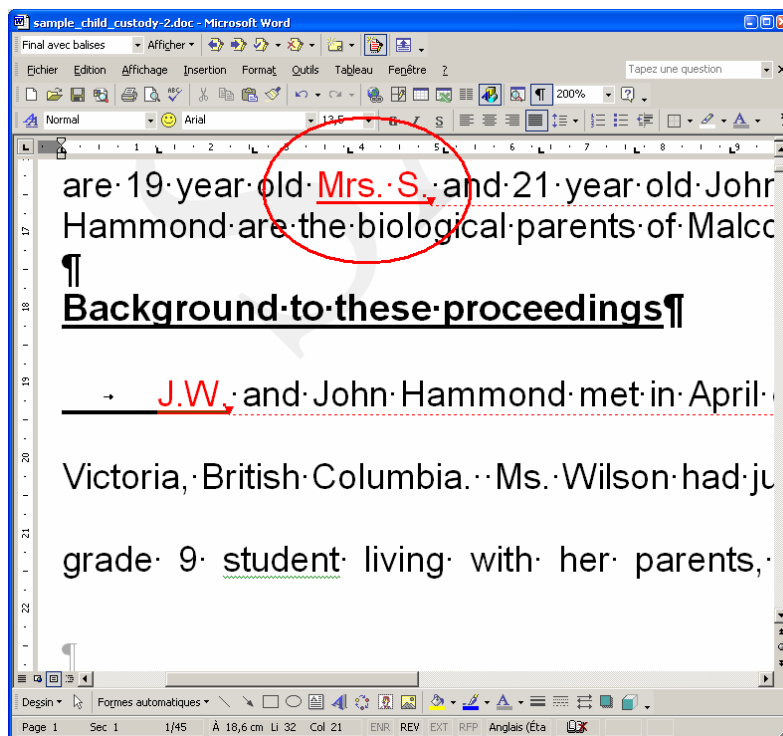
If a judge's name has been replaced by accident, the user can repeat the replacement, this time with the full name instead of the initials. In such a case, the mistakenly inserted initials will be replaced by the original full name.



[Figure 8: Undoing a replacement]



Single replacements of a term can be modified by rejecting the specific single change in the document resulting from the systematic replacement. (Fig. 9)



If all but one replacement of **Mrs. Smith** is relevant, the user can reject the single erroneous replacement using the MS Word Track Changes features.



[Figure 9: Rejecting a single change]

Saving the replacements

All the replacements appear in the document as changes requiring approval. The changes can be accepted in their entirety by clicking on . Similarly, the changes can be entirely rejected by clicking on .

Using  or  affects all the changes made to the document regardless of the selections in the replacement list.

When the user considers that the changes are final, the document must be saved. It must be noted however that saving the document will not save the changes made to the NOME replacement suggestions list for the specific session. These changes, such as the inclusion of a term for the session or the removal of a term for a session, will be lost as the user quits the session.

In order to avoid loss of data, the user should save the document periodically during the editing process.

Refining NOME's basic detection logic

NOME's detection logic can be refined by modifying its inclusion, title and exclusion lists. Customizing these lists is strongly encouraged since it can significantly improve NOME's performances in detecting relevant proper names.

Inclusion list

The inclusion list contains terms that NOME should always detect as relevant even though they do not satisfy the basic detection logic requirements.

Terms such as ***social insurance number***, ***born on***, and ***bank account*** can be added to the inclusion list.

Titles list

Normally, NOME's basic logic avoids the detection of any capitalized term preceded by a dot, in order to ignore words that are capitalized only because they begin a sentence. However, titles of persons often end with a dot. NOME detects capitalized terms preceded by titles such as "Ms." or "Dr." only if they are included in the titles list.

According to NOME's basic logic, the term ***Mr. Marlowe*** would be ignored, but since the title ***Mr.*** is included in the titles list, it will be detected as a proper name.

Exclusion list

The exclusion list contains terms that will always be ignored by NOME, despite its basic logic. The exclusion list should be managed prudently to avoid undesired non-detection of relevant terms.

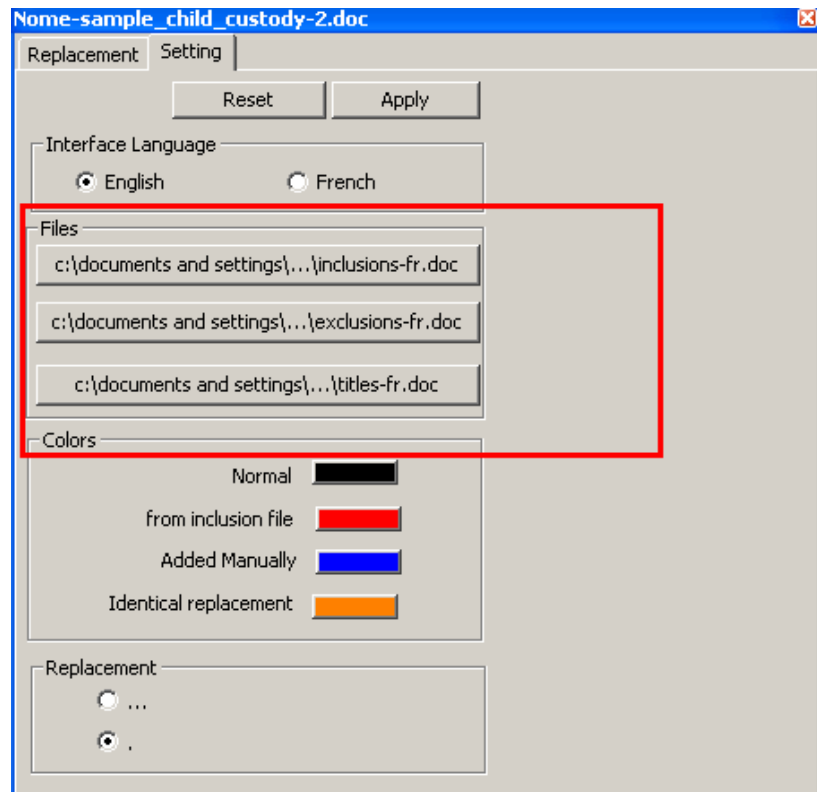
In the context of a court, terms like ***Provincial Court***, ***Criminal Code*** may be added to the exclusion list in order to avoid their detection as a person's proper name.

Changing and creating lists

By default, NOME provides one inclusion, titles and exclusion list per language. It may be appropriate for a user to change the content of these lists or to create different lists according to the context of use. Different inclusion or exclusion lists can be created, for example, by court or by subject matter.

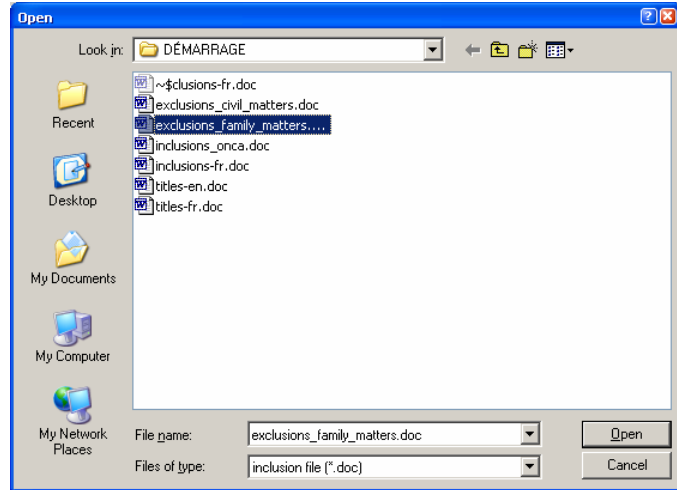
The user can create different inclusion lists that can be selected when needed, such as *inclusion_civil.doc*, *inclusion_family.doc* or *inclusion_onca.doc*.

The user can select and add or remove different list files by clicking on the inclusion list buttons on the **Setting** menu of the NOME interface. NOME will automatically apply the last selected inclusion, titles and exclusion lists to the subsequent session or reanalysis of the document. (Fig. 10)



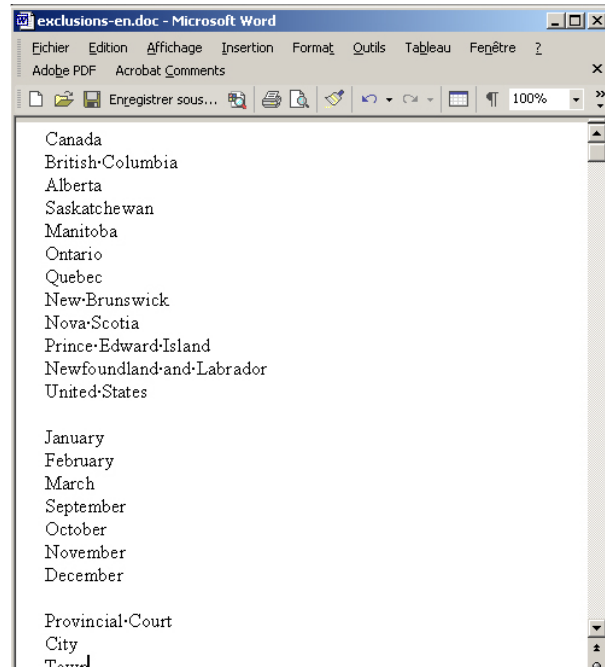
[Figure 10: The Setting menu]

These lists are MS Word files that can be stored anywhere on the user's hard drive. (Fig. 11)



[Figure 11: Inclusion, Exclusion and Titles lists]

The lists are quite easy to change or to create since they are regular MS Word documents. The only requirement is that each term should be placed on a separate line. (Fig. 12)



[Figure 12: Content of an exclusion list]

Other settings

Besides management of the inclusion, titles and exclusion lists, the **Settings** menu offers the following possibilities:

- Choice of interface language;
- Choice of colors;
- Choice of replacement method.

In order to save the settings for the interface language, colors and replacement method for subsequent sessions, the user must click on the “Apply” button. Otherwise, these settings will apply to the current session only.

Recommended workflow

- Launch NOME;
- Purge obviously irrelevant terms from the list of replacement suggestions;
- Make decisions about whether to replace or ignore other terms by browsing through their occurrences in the text of the document;
- Tag the terms to be ignored and remove them from the list;
- Tag the terms to be replaced and check the replacement initials;
- Make the replacements;
- Check the result and accept the modifications made to the document;
- Save the anonymized copy of the document.