RASH in DOCX

Documentation – Version 0.1, October 1, 2016

# Abstract

This document describe how to use Microsoft Word for writing scientific documents that can be easily converted in the Research Article in Simplified HTML (RASH) format. RASH is a markup language defined as a subset of HTML+RDF for writing scientific articles – its full documentation is available at <https://rawgit.com/essepuntato/rash/master/documentation/index.html>. Note that this section is identified as abstract section because of the particular heading we have used (i.e., all sections having heading either Abstract or Summary are intended as abstract sections).

# Introduction

This paper introduces how to use Microsoft Word in order to write scientific articles that are easily convertible into RASH, the Research Articles in Simplified HTML format 1. RASH is a markup language that restricts the use of HTML elements to only 25 elements for writing academic research articles and that allows us to include also RDFa annotations within any element of the language.

The whole conversion approach has been proposed in order to provide authors using Microsoft Word an easy way to write articles without caring about the particular HTML tags to use for defining particular blocks of text and metadata. In addition, this is a meta-document (and, thus, a sort of template for scientific documents), since it uses all the admissible functionalities of Microsoft Word that still allow one to easily converted it into RASH by using the tools we introduce at the end of the document. This document basically explains how to use Microsoft Word default styles and tools for expressing the various elements that form a scientific paper. It is worth mentioning that, in order to have the best conversion into RASH, **only the default styles (with no modifications) and the listed tools should be used**. In particular, please do not care how the various elements will be shown in the DOCX document, since Microsoft Word should be used only for writing the content of the paper rather than presenting such content. All the presentational layout of the text will be handled by other external tools after converting the paper into RASH.

The rest of the document is organized as follows. In Metadata we explain how to specify the metadata of the article, such as titles, authors, affiliations, keywords and ACM categories. In Textual elements we introduce all the elements responsible for the definition of the actual text of documents, i.e, paragraphs and inlines. In Sections we present the hierarchical structures for organising the text, e.g., the sections. In Boxes for figures, tables, listings and formulas we show how to create figures (both inlines and as separate blocks), tables and formulas (both inlines and as separate blocks). In References section we introduce how to define the bibliographic reference of a paper. In Footnotes we show how to define footnotes related to a document. In Referring things we explain how to create references to all the other objects defined with an article, for instance in-text reference pointers, footnote pointers, etc. Finally, in Conversion into RASH we present the tools for converting DOCX documents into RASH.

# Metadata

Each scholarly article comes with particular metadata that characterize the document, i.e., the title, the authors and (optionally) the keywords and ACM categories of the paper.

## Title and subtitle

We can specify the title of the paper by using the style Title, and the subtitle by using the style Subtitle. The style panel is shown when clicking on the button  or the button  on the Microsoft Word toolbar. In Figure 1 the “Style and Formatting” panel is shown. You can select the new style for the paragraph currently containing the cursor by clicking on the style name of interest (in the figure, the style Normal is selected). In order to have all the styles available for selection, it is important to choose the All Styles option from the drop-down menu in the bottom part of the panel.

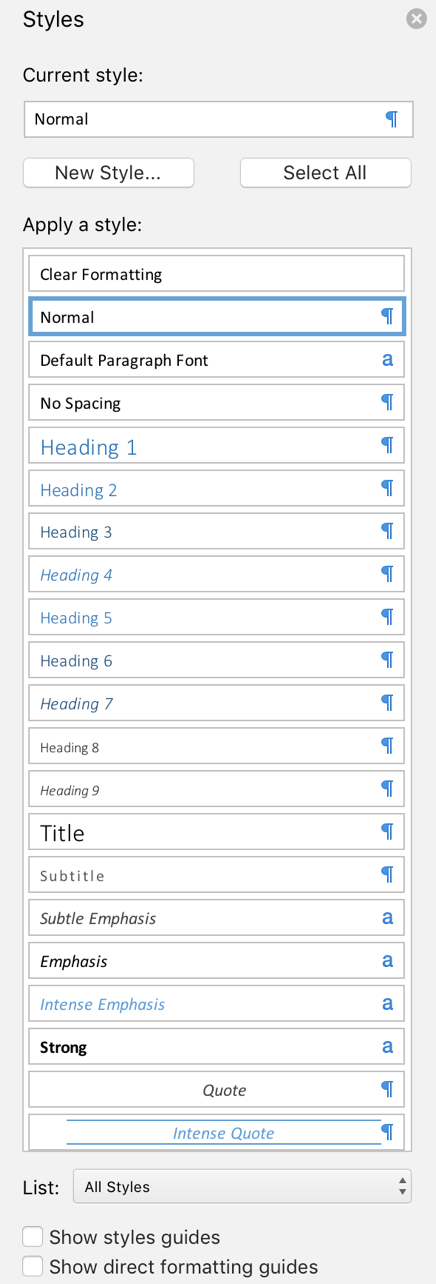


Figure 1 The style and formatting pane, which allows one to select any style for the block of text in the document.

## Authors and affiliations

The authors of a document and all their related information (i.e., name, email, and affiliation) can be defined by using a particular user-defined metadata of the document containing a string defining her name, email, and affiliation. In order to do that, one has to open the properties of the current document by clicking on “File” and then “Properties...”, as shown in Figure 2.

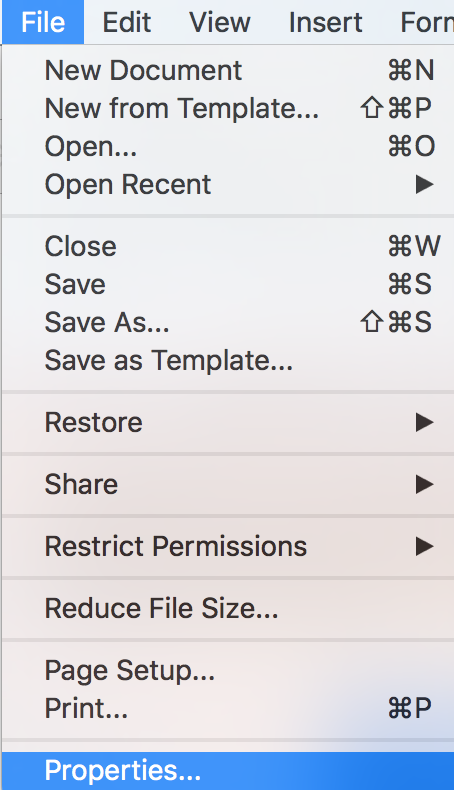


Figure 2: The file menu, which allows the access to the properties of the document.

In the “Custom” panel, we can add additional metadata to the document by clicking on the button “Add” – see Figure 3. The information about an author must be specified by using the property starting with *Author* (e.g., Author\_1) having type *Text*, and by specifying a particular string of tokens separated by “--“ as a value of such property. For instance, for specifying the author “Alberto Nicoletti” having email “alberto.nicoletti@studio.unibo.it” and affiliation “Department of Computer Science and Engineering, University of Bologna, Bologna, Italy”, we should use the following string as value of the property:

Alberto Nicoletti -- alberto.nicoletti@studio.unibo.it -- Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

Once added all the authors, we can proceed by clicking “OK”. In case more than one authors must be specified, one can simply add additional “Author” properties with different indexes, e.g., *Author\_2*, *Author\_3*, *Author\_4*, as shown in Figure 3.

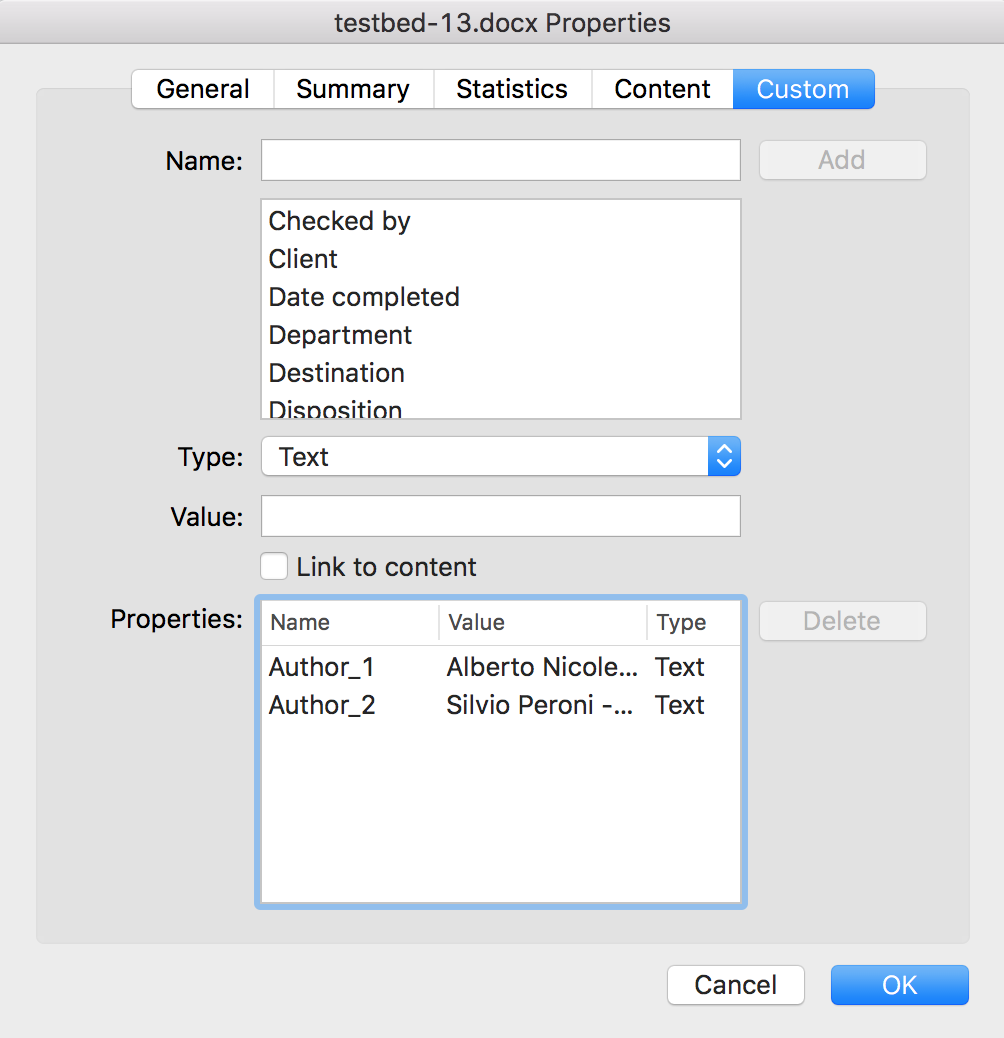


Figure 3, The panel for adding custom properties with keywords and categories specified.

## Keywords and ACM categories

In addition to the aforementioned data, it is possible to use the custom properties to specify other article-related data, such as the ACM subject categories and free text keyword. In particular:

* the property *Keywords* (having type *Text*) can be used to add keywords related with the article (use “--“ as separator);
* the property *Categories* (having type *Text*) can be used to add an [ACM subject category](http://www.acm.org/about/class/class/2012) related with the article (use “--“ as separator).

An example of how to create all these information is shown in Figure 4.

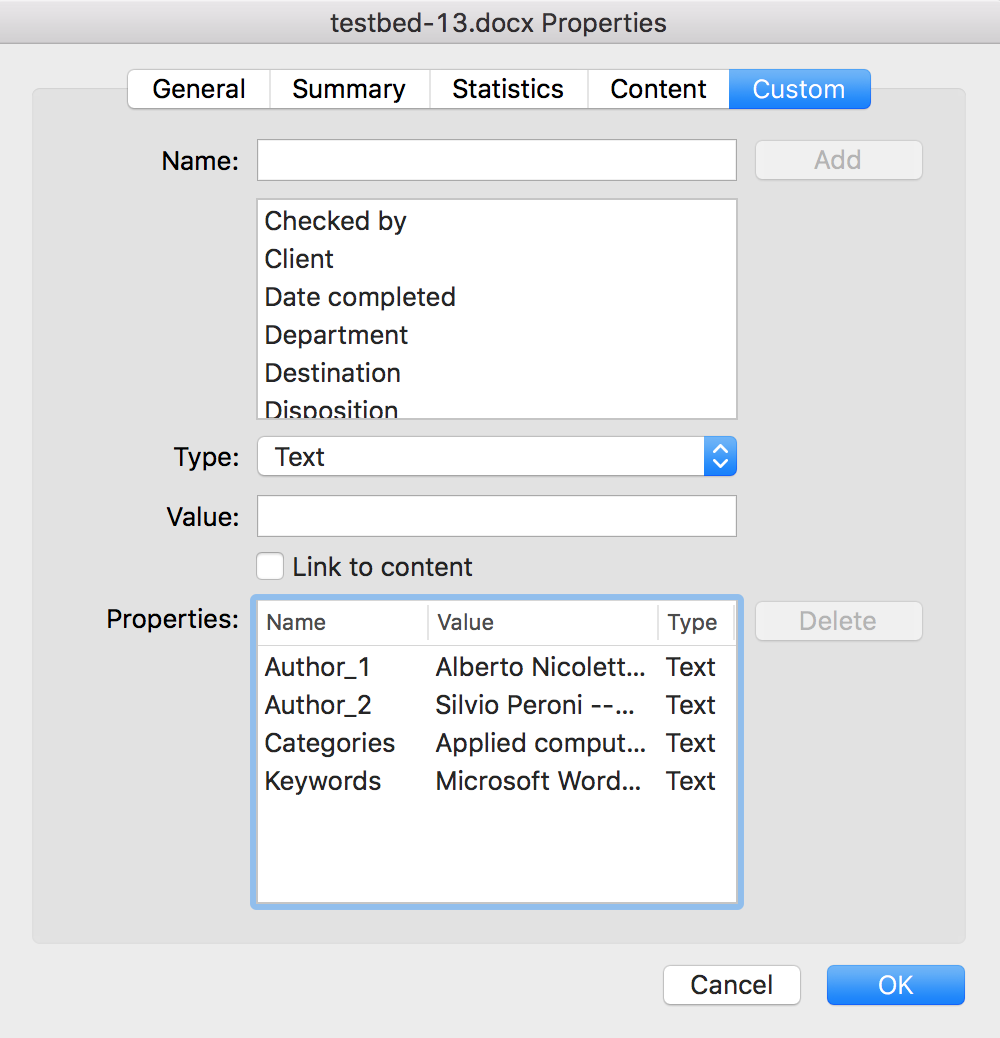


Figure 4: The panel for adding custom properties with keywords and categories specified.

# Textual elements

The textual elements for describing the content of the paper can be split in two main categories. On the one hand, there are those that allow us to organise the content of a document in textual blocks, i.e., the paragraphs and other elements containing text. On the other hand, there are elements that allow us to associate particular functions to portions of text, i.e., the inline elements.

## Paragraphs

First of all, there are the paragraphs, i.e., “a self-contained unit of a discourse in writing dealing with a particular point or idea” 2. A paragraph is the simplest block structure containing text and, among the blocks, the only kind of element in the body where you can use the text.

There exist three different kinds of paragraphs: normal (as this one), paragraphs containing codes and paragraphs with quotes.

The normal paragraphs are defined by means of the standard style *Normal*, which is usually selected as default by Microsoft Word.

Then we have paragraphs containing source code. They are defined by using the style *HTML Preformatted* available in the “Style pane”, as shown as follows.

This is a paragraph containing codes

This is another one

Finally, it is possible to also describe blocks of quotations by using the style *Citation* as follows:

This is a quoted paragraph.

## Lists

The lists are self-contained blocks that must be used outside paragraphs and that allow us to structure text as a sequence of item. The buttons /Users/alberto/Desktop/Screen Shot 2016-10-02 at 12.09.53.png in the toolbar are used to define:

* unordered lists (like this one), and
* ordered lists (like the following one).

It is possible to include more than one paragraph within each list item as follows:

1. This is the first item of the list.

This item includes two paragraphs.

1. This is the second item of the list.

## Inline elements

In addition to blocks, it is possible to describe small parts of a text with inline elements defining particular functions on such text, such as emphasis, inline quotation, super- and sub- scripts, and the like.

We can use two different kinds of emphases, the normal one and the **strong one**, created by using the button/Users/alberto/Desktop/Screen Shot 2016-10-02 at 12.21.26.png (CTRL+I on Windows or CMD+I on Mac) and the button /Users/alberto/Desktop/Screen Shot 2016-10-02 at 12.21.40.png (CTRL+B on Windows or CMD+B on Mac) respectively. Inline quotations, e.g., “Alice was beginning to get very tired of sitting by her sister on the bank”, are possible by simply writing the opening and closing quotations. While superscripts and subscripts are defined using the buttons /Users/alberto/Desktop/Screen Shot 2016-10-02 at 12.26.00.png (CTRL+SHIFT+= on Windows or CMD+SHIFT+= on MAC) and /Users/alberto/Desktop/Screen Shot 2016-10-02 at 12.25.52.png (CTRL+= on Windows or CMD+= on MAC) respectively.

This paragraph contains a simple example of inline code, created by means of the default style for characters *HTML Code*. Alternatively, it is possible to directly use the Courier font style and the Courier New font style for having the same effect.

It is also possible to insert images and formulas as part of the text inside a paragraph. For images, one should click on  button available in the “Insert toolbar”, or just drag-and-drop it into the document. In addition, in order to place the image in the right position within a paragraph, it is important make sure it is in line with the text (it is default behaviour). This is possible by right-clicking on the image and then selecting the option In “Line with Text” In in the menu “Wrap Text”, as shown in Figure 5.

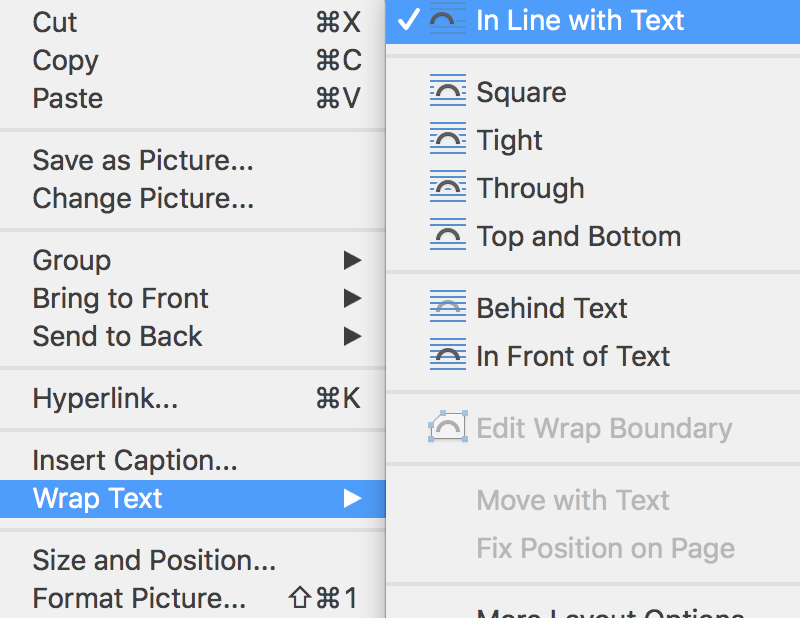
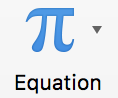


Figure 5The panel that allows one to set the image in line with text.

For formulas, one click the “Equation button”  in the “Insert toolbar”. Now one can insert the formula using the tools contained in the “Equation toolbar” (as shown in Figure 6) in order to create a formula such as .

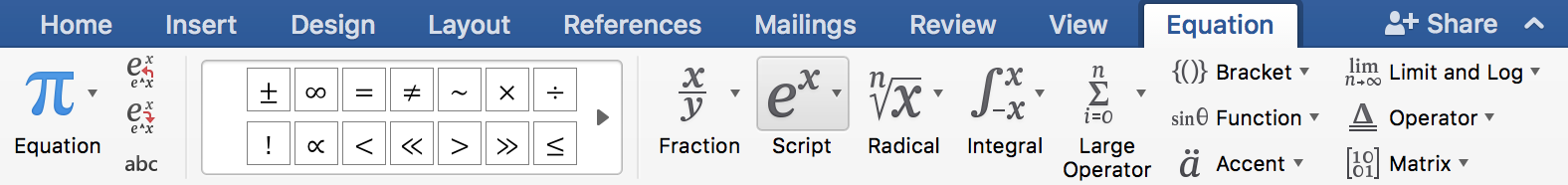


Figure 6The equation toolbox.

Finally, the [hyperlinks](https://github.com/essepuntato/rash) can be created by selecting the text, right-clicking on it and selecting “Hyperlink...” in the context menu as shown in Figure 6 or by clicking “Hyperlink” in “Links”, contained in the “Insert toolbar”, as shown in Figure 7.

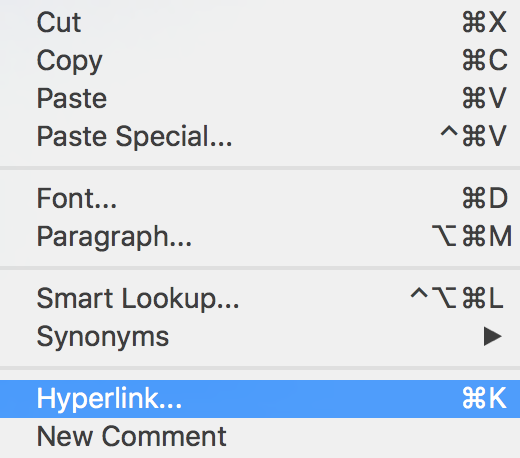


Figure 7The panel that allows one to create an hyperlink from a selected text.

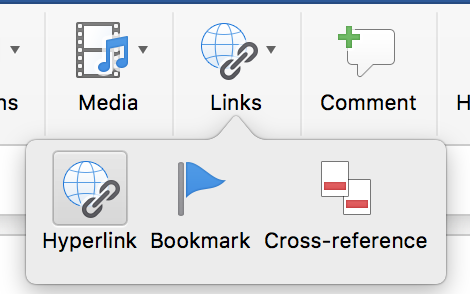


Figure 8The button that allows one to add an hyperlink.

# Sections

Sections are pure structural constructs that allow one to organize the various textual blocks of a document according to appropriate hierarchies. In Microsoft Word, any new section is implicitly defined by specifying a new heading by means of the appropriate style: *Heading 1*, *Heading 2* and *Heading 3*.

The sequence of headings of different importance implicitly describes the organization into sections and subsections of the document. Thus, main sections will be added by means of the style *Heading 1*, subsections of those ones will be created by using the style *Heading 2*, and sub-subsections of the latter ones will be created by using the style *Heading 3*.

# Boxes for figures, tables, listings and formulas

In a scientific article, figures, tables and formulas as usually contained within appropriate blocks accompanied by a caption that allow one to refer to such items easily within the text. In Microsoft Word, all these boxes are defined easily by inserting the object of interest and then by associating a caption to it.

## Figure boxes

We can define a figure box containing an image with a caption by adding a new picture in an empty paragraphs (as introduced in Inline elements) and the by right-clicking on the added picture and selecting the field “Insert Caption…”, as shown in Figure 8.

Then one can simply add the text of the caption by using the panel “Caption” shown in Figure 9.

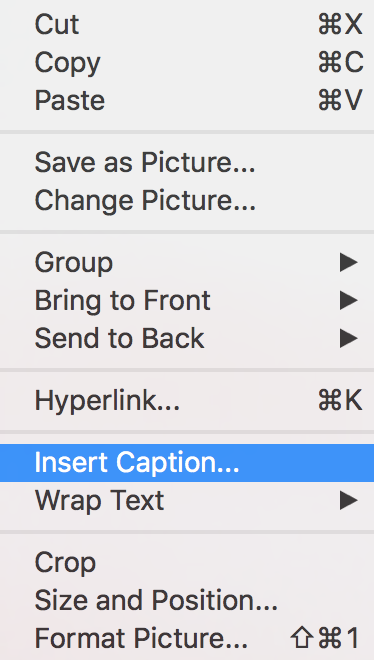


Figure 9The panel for adding captions to objects.

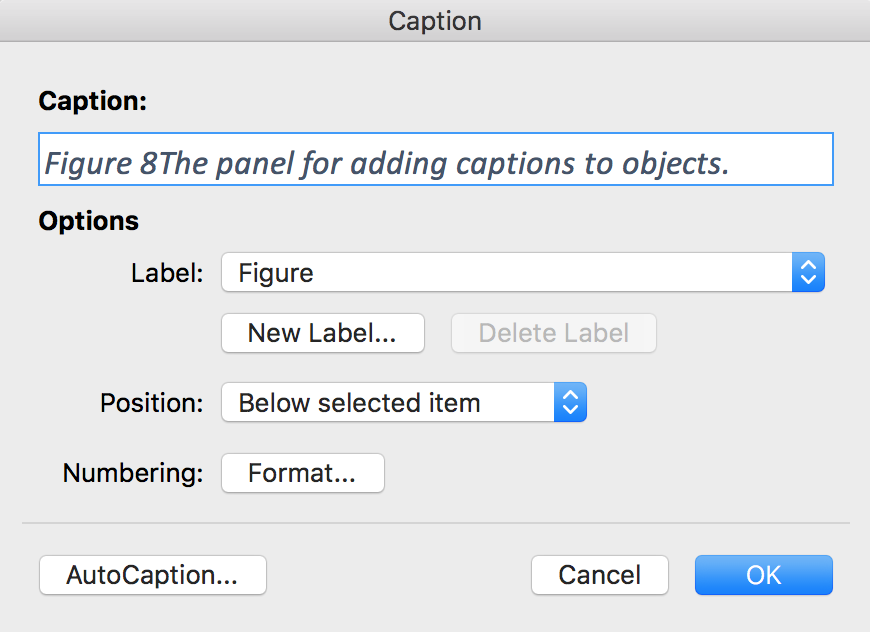


Figure 10The panel for adding the text of the caption to the selected object.

## Table boxes

We can define a table box containing a table with a caption by clicking on the button /Users/alberto/Desktop/Screen Shot 2016-10-02 at 17.39.48.png located in the “Insert toolbar”, when the cursor is on an empty paragraph and then selecting how many rows and columns are needed for the table. The caption can be added by selecting the whole table, right-clicking on it and selecting the field “Insert Caption...”, as also shown for figures in Figure boxes.

As shown in Table 1, a table can have an header, in order to have it, we must select the option “Header Row” in the “Table Design toolbox”, as shown in Figure 10.

|  |  |
| --- | --- |
| Header 1 | Header 2 |
| Cell 11 | Cell 12 |
| Cell 21 | Cell 22 |

Table 1An example of Table with headers

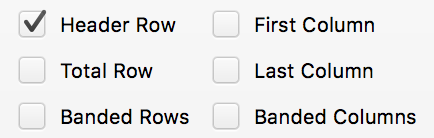


Figure 11The option to select in order to have an heading row in a table.

## Listing boxes

We can define a listing box containing text with a caption by adding a new “Text box” in an empty paragraphs as shown in Figure 10, and then drawing a text box of the preferred size. Then the caption can be added by right-clicking on the added Text box and selecting the field “Insert Caption...”, similar to what has been shown in Figure 8.

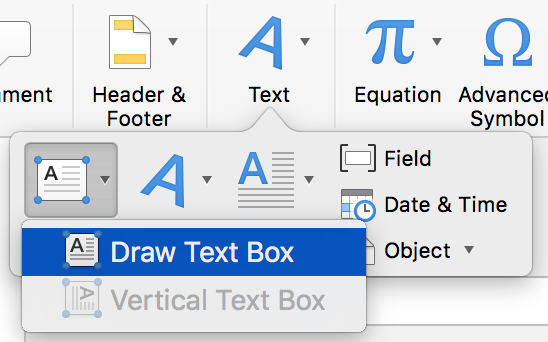


Figure 12This is the button used to draw a text box, in order to create a listing box.

As shown in Listing 1, the content of the frame should be only textual, it is possible to use multiple paragraph and does not matter what particular style for the paragraph one want to use. However, as a sort of guideline, we strongly suggest to use the Preformatted Text available in the Style and Formatting panel (CTRL+T for Windows users, CMD+T for Apple users) as also shown in Paragraphs.

This is a paragraph to put in listings.

It is not important which font you use here, even if it is common to use the HTML Preformatted style.

Text 1This is the caption of the code listing.

## Formula boxes

We can define a formula box (useful for allowing formulas to be referenced within the text) containing a formula with an empty caption by adding a new formula in an empty paragraphs (as introduced in Inline elements) and then by selecting the added formula and selecting the field “Caption...” in the “Insert” menu as shown for figures in Figure 11 without specifying any text associated.

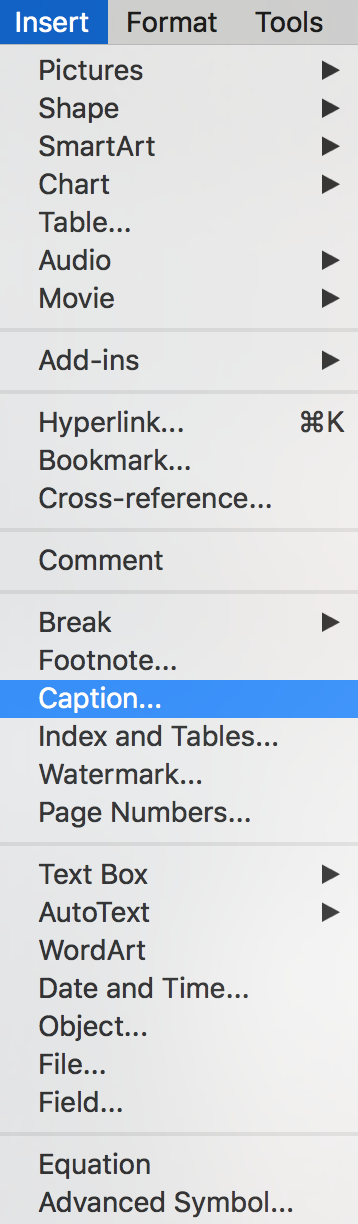


Figure 13The button which has to be clicked to insert a caption on a formula.

For instance the formula introduced in Inline elements can be expressed in a formula box as shown in Equation 1.

Equation 1

# References section

All the references are organized in an ordered list within a special section defined by the actual content of the heading of such section. In particular, all sections having heading either “References” or “Bibliography” are intended as reference sections.

This special section must have an heading and **only one** ordered list containing the full text of a bibliographic reference for each list item. Usually, it is the last section of the document. It is worth mentioning that the user can avoid to take care of the order the various references are listed in the reference list, since they will be reordered automatically in the RASH document obtained from the conversion of the original DOCX document.

# Footnotes

All the footnotes like[[1]](#footnote-1) can be defined by opening the menu “Insert”, selecting the option “Footnote...”, and by clicking on the button “Insert” in the footnote panel shown in Figure 11, taking care that the location selected is “Footnote”.

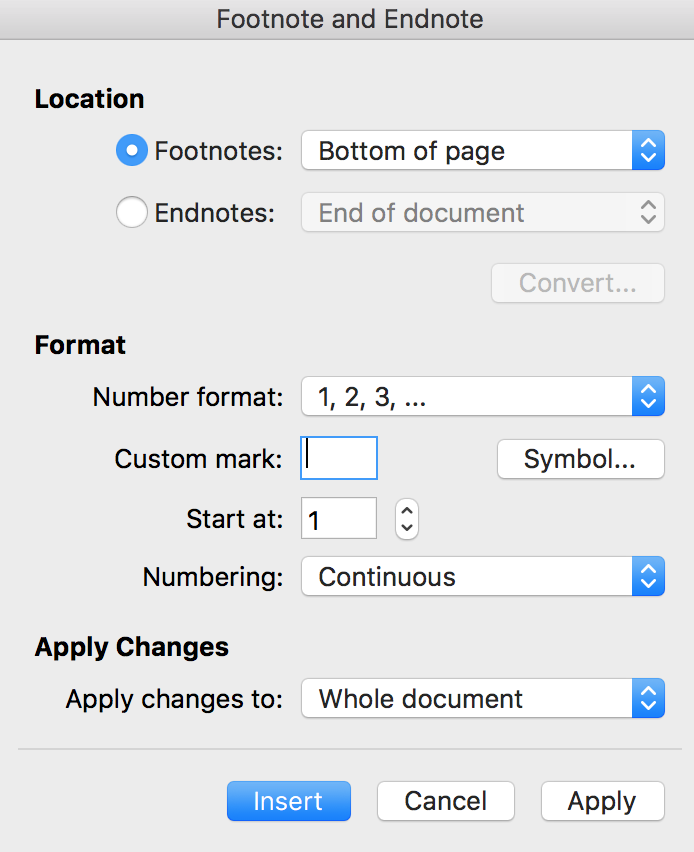


Figure 14The panel for creating new footnotes.

# Referring things

All the sections (e.g., Introduction), figures (e.g., Figure 1), tables (e.g., Table 1), listings (e.g., Text 1), formulas (e.g., Equation 1), bibliographic references (e.g., 2), footnotes (e.g., 1) can be referenced within the text by using the Microsoft Word cross-reference tool available by selecting the menu “Insert” and then clicking on “Cross-reference...” The panel that will be opened (shown in Figure 12), allows us to specify to what particular object we want to refer to (section “Reference type”), which object of the selected type should be referenced (section “For which caption”), and which referencing style should be used within the document (section “Insert reference to”).

While different option for including referencing in the document are possible, we suggest to use specific referencing style according to the object that is referenced. Our suggestion are described in Table 2– of course the user can choose to adopt the preferred referencing style that differs with our suggestions.

|  |  |  |
| --- | --- | --- |
| Object | Type | Referencing style |
| Section | Heading | Heading text |
| Figure | Figure | Only label and number |
| Table | Table | Only label and number |
| Listing | Text | Only label and number |
| Formula | Equation | Only label and number |
| Bibliographic reference | Numbered Item | Paragraph number (full-context) |
| Footnote | Footnote | Footnote number (formatted) |

Table 2The objects that can be referenced in the document (and that will be correctly converted in RASH) with their type and referencing style suggested.

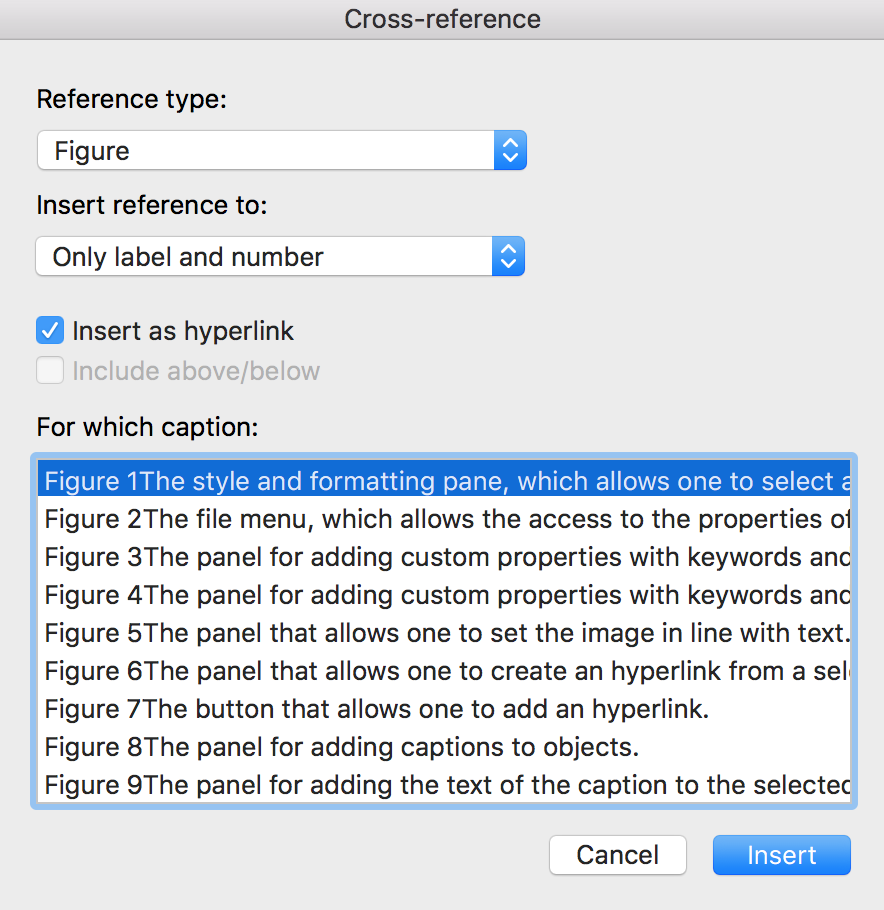


Figure 15The panel for adding cross-references to objects of the document.

# Conversion into RASH

The conversion of DOCX documents, such as this one, into RASH is implemented by using a particular XSLT 2.0 document we developed for that purpose. The XSLT file is available online at <http://github.com/essepuntato/rash/xslt/from-docx.xsl>. In addition, we have also released a simple Java application, downloadable at <http://github.com/essepuntato/rash/tools/docx2rash> called DOCX2RASH, that includes such XSLT file and that allows one to automatically convert an ODT document into RASH. To run the application you need to run the following command:

java –jar docx2rash.jar –i <input-docx-file> -o <new-output-directory>

# Acknowledgements

The acknowledgements of a document are organized in a special section defined by the actual content of the heading of such section. In particular, all sections having heading either Acknowledgement or Acknowledgements are intended as acknowledgements sections. Acknowledgements are usually put just after the section concluding the document (i.e., Conversion into RASH in this case) and before the reference section.

# References

1. Peroni, S. (2015). RASH: Research Articles in Simplified HTML – Documentation. Version 0.4, October 22, 2015. <https://rawgit.com/essepuntato/rash/master/documentation/index.html>
2. Wikipedia entry Paragraph. <https://en.wikipedia.org/wiki/Paragraph> (last visited, October 30, 2015).

1. This is the first footnote of the document. [↑](#footnote-ref-1)