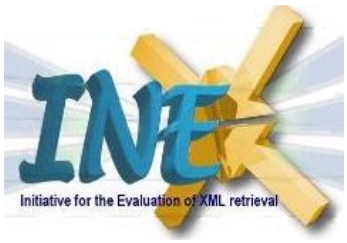


INEX 2007 Relevance Assessment Guide

Mounia Lalmas and Benjamin Piwowarski[+]

1. Introduction

During the retrieval runs, participating organisations evaluated the 130 INEX 2007 topics (CO+S) against the Wikipedia document collection and produced a list (or set) of document components (XML elements[1]) as their retrieval results for each topic. The top 1500 components in a topic's retrieval results were then submitted to INEX. The submissions received from the different participating groups have now been pooled and redistributed to the participating groups (to the topic authors whenever possible) for relevance assessment. Note that the assessment of a given topic should not be regarded as a group task, but should be provided by one person only (e.g. by the topic author or the assigned assessor).

The aim of this guide is to outline the process of providing relevance assessments for the INEX 2007 test collection. This requires first a definition of relevance (Section 2), followed by details of how to assess (Section 3). Finally, we describe the on-line relevance assessment system that should be used to record your assessments (Section 4).

2. Relevance in INEX

Relevance in INEX is defined according to the notion of **specificity**, which describes the extent to which the document component focuses on the topic of request. This definition was adopted after a number of studies that showed that in terms of retrieval effectiveness, the same conclusions could be in most cases generated from using the specificity dimension of relevance compared to using more complex definitions. Up to INEX 2005, relevance was defined according to two dimensions, specificity and exhaustivity. The latter describes the extent to which the document component discusses the topic of request. This year (as for 2006), only the specificity dimension is used. Its measuring is based on the highlighting procedure used since INEX 2005. The main advantage of this highlighting approach is the specificity of any (partially highlighted) elements can be calculated automatically as some function of the contained relevant and irrelevant content (e.g. in the simplest case as the ratio of relevant content to all content, measured in number of words or characters).

3. How to assess

The assessment process is to be done as follows. Assessors highlight text fragments that contain only relevant information. It is important that only purely relevant information fragments get highlighted. To decide which text to highlight, you should skim-read the whole article and identify any relevant information as you go along. The on-line system can assist you in this task by highlighting keywords (that are chosen using the interface) and pool elements (elements retrieved by participating systems) within the article (see Section 5). If you highlight any part of a document, the document is considered relevant. For any such document, you should also select a so-called "best entry point" (BEP) of the document.

During the relevance assessment of a given topic, all parts of the topic specification should be consulted in the following order of priority: narrative, topic description, and topic title. The narrative should be treated **as the most authoritative description of the user's information need**, and hence it serves as the main point of reference against which relevance should be assessed. In case there is conflicting information between the narrative and other parts of a topic, the information contained in the narrative is decisive. *Note that it is not because that a term listed within the topic is not present in an element that the element is not relevant. Similarly, the presence of topic terms does not imply its relevance.* It may be that a component contains some or maybe all the terms, but is irrelevant to the

[+]Based on a prior guidelines authored by M. Lalmas, B. Piwowarski and G. Kazai

[1] The terms document component and XML element are used interchangeably.

topic of the request. Also, there may be components that contain none of the terms yet are relevant to the topic.

For the CO+S, the topic titles (may) contain structural constraints in the form of XPath expressions. These structural conditions should be ignored during your assessment. This means that you should assess the elements returned for a CO+S topic as whether they satisfy your information need (as specified by the topic) **with respect to the content criterion only**.

You should judge each text fragment on its own merit. That is, a text fragment is still relevant even if it is the twentieth you have seen with the same information. It is imperative that you maintain consistency in your judgement during assessment. Referring to the topic text from time to time will help you maintain judgement consistency.

4. Using the on-line assessment system (X-Rai)

There is an on-line relevance assessment system (XML Retrieval Assessment Interface) provided at:

https://inex.lip6.fr/2007/adhoc

which allows you to view the pooled result set of the topics assigned to you for assessment, to browse the Wikipedia document collection and to record your assessments. Use your INEX username and password to access this system.

The assessment tool works with opera and recent "gecko" browsers: we highly recommend you to use Opera (version 8 or up only; version 9 is recommended) available at http://www.opera.com. Other compatible browsers are:

- **Mozilla** (version 1.7 or up) at http://www.mozilla.org/products/firefox/.
- **Firefox** (version 1 and up) at http://www.mozilla.org/products/mozilla1.x/.

Note that **JavaScript must be enabled** for the assessment tool to work and that **the assessment tool is not compatible with Internet Explorer. Any bug report should be submitted using the project homepage (**https://developer.berlios.de/projects/x-rai/**) using the link in the "Links" menu of the interface (Figure 1).**

4.1. Home page

After logging in, you will be presented with the Home page (see Figure 1) listing the topic ID numbers of the topics assigned to you for assessment (under the title "Choose a pool"). This page can always be reached by clicking on the **"X-Rai"** link of the menu bar on any subsequent pages.

Each X-Rai page is composed of the following components:

- The menu bar, which is itself composed of four parts:
    1. The login name (e.g. "demo" in Figure 1),
    2. A list of menu items, which can be accessed by holding the mouse over the menu label (e.g. "**Links**" in Figure 1.),
    3. The location within X-Rai, where each location step is a hyperlink (in Figure 1, we are at the root of the web site, so the only component of the location is "**X-Rai**", which is a link to the home page),
    4. The menu bar may also contain a number of icons (displayed on the right hand side, see Figure 2a). Click on one of these icons to display (or hide):
        - Information about X-Rai.
        - Toggle the help displayed when holding the mouse over icons or hyperlinks
- The main window.
- An optional status bar (see Figure 4), displayed only when assessing a pool, i.e. in pool, sub-collection or article view (see relevant sections below) appears at the bottom of the window and shows the number of unknown assessments you have to judge before completing assessing the document (in Figure 4, there is only one unknown assessment).
- In the status bar, three arrows (, and) may be used to navigate quickly between the elements to be assessed. You may also use the shortcut keys of 1 (left), 2 (up) and 3 (right). The up arrow enables you to move to a level up in the hierarchy, e.g. from an article or a collection part to its innermost enclosing part of the collection (you move in the opposite direction by

selecting a sub-collection or an article). The left arrow can be used to go to the previous element to be assessed, while the right arrow to go to the next element to be assessed.

The on-line assessment system provides three main views (Sections 4.2 to 4.4):

1. Pool view,
2. Sub-collection view, and
3. Article view

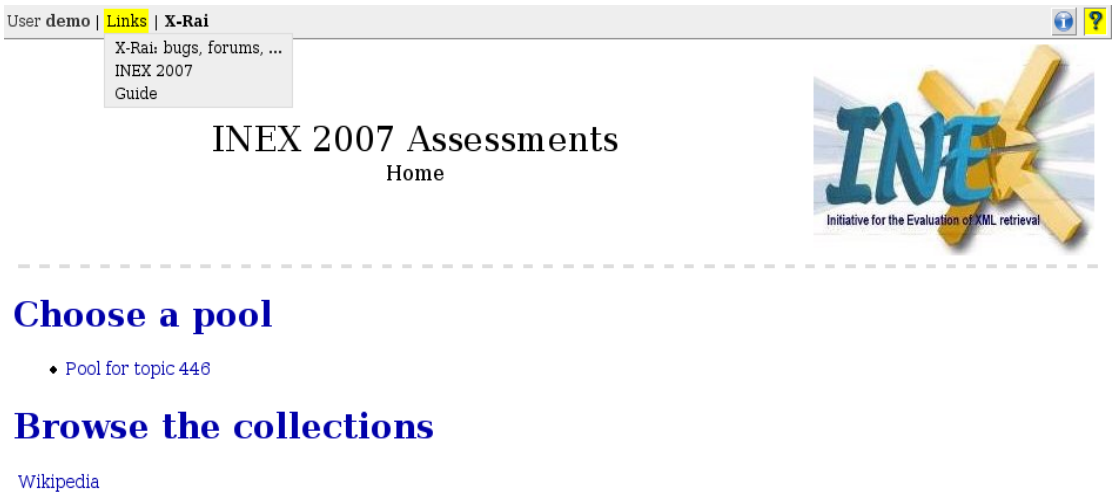


Figure 2: Home page and menu bar

In the "**Links**" menu

- **INEX 2007**: link to the official INEX web site.
- **X-Rai project:** link to the development web site of X-Rai where you can submit bug reports or/and feature requests.
- **Guide:** the latest version of this assessment guide.

4.2. Pool view

Clicking on a topic ID will display the Pool main page for that topic (see Figure 2a).

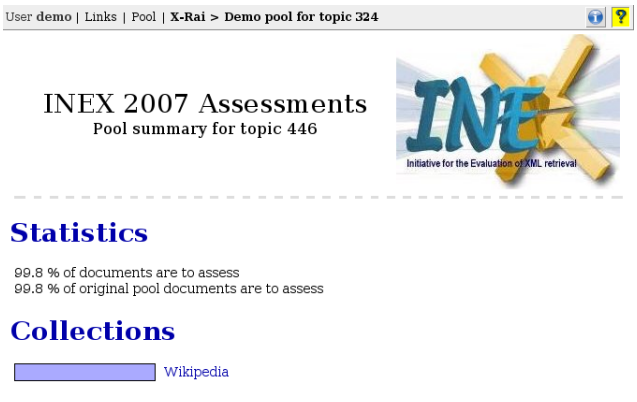


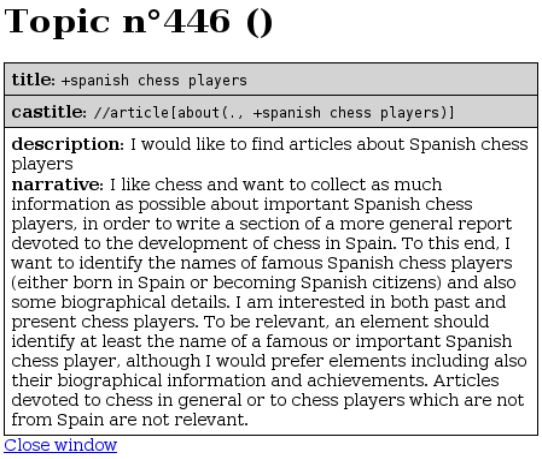
Figure 2a. Pool view

Figure 2b. Pool menu:

Topic: displays topic statement.
Keywords: to manipulate list of words and phrases to highlight.
Support: to change the way support elements are displayed.
History: show the last viewed documents

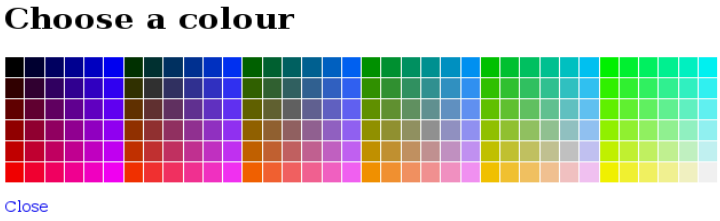
Here, a new menu item, "**Pool**", appears on the menu bar at the top of the window. This menu item will remain whenever you are viewing a pool related page.

Within the "Pool" menu (Figure 2b), with the "**Topic**" submenu item you can display the topic statement in a popup window. This is useful as it allows you to refer to the topic text at any time during your assessment. An example of the topic popup window is given below:

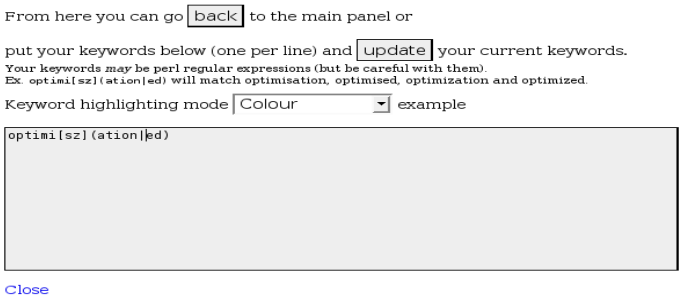


The "**Keywords**" submenu item allows you to access a feature, where you can specify a list of words or phrases to be highlighted when viewing the contents of an article during assessment. These cue words or phrases can help you in locating potentially relevant texts within an article and may aid you in speeding up your assessment (so add as many relevant cue words as you can think of!). You may edit, add to or delete from your list of keywords at any time during your assessment (remember, however, to refresh the currently assessed article to reflect the changes).

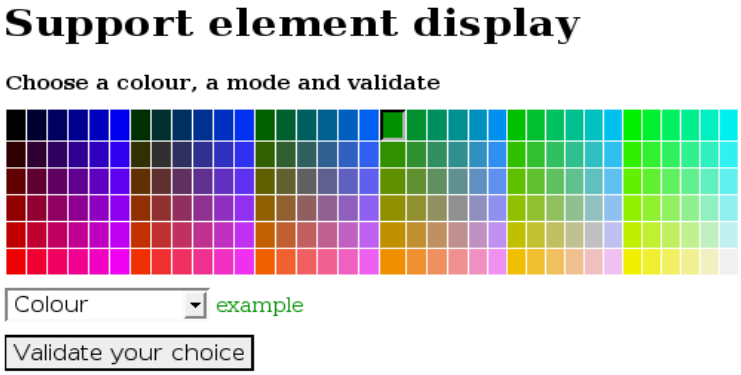
You may also specify the preferred highlighting colour for each and every keyword. After selecting the "**Keywords**" menu item, a popup window will appear showing a table of coloured cells. A border surrounding a cell signifies a colour that is already used for highlighting some keywords. Move the mouse over a coloured cell to display the list of keywords that will be highlighted in that colour. To edit the list of words or phrases for a given colour, click on the cell of your choice.



You will be prompted to enter a list of words or phrases (one per line) to highlight (see figure below). You can choose three different highlighting modes using the drop-down menu: using coloured fonts, drawing a border around the phrase or using a background colour. You can use different highlighting modes with the same colour. To edit the list of words for a given mode, select the highlighting mode in the drop-down menu. You can then edit the list in the text area below. Note that the words or phrases you specify will be matched against the text in the assessed documents in their exact form, *i.e.* no stemming is performed. Your keywords *may* be perl regular expressions (but be careful with them). For example, `optimi[sz](ation|ed)` will match optimisation, optimised, optimization and optimized.



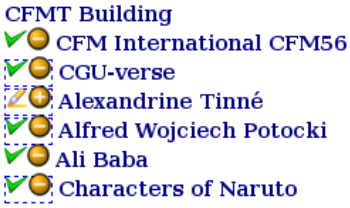
The "**Support**" item allows you to change the way support elements (i.e. elements retrieved by the participating systems) are displayed. When selecting this menu item, a pop-window (shown below) appears and allow you to change the colour (clicking on a colour) and the mode (background, font colour, or border, by selecting an item in the drop-down menu) of the highlighting. An example of the support element display is shown at the right of the mode selection.



The "**History**" item allows you to access the list of last viewed documents, which can be useful if you want to go back to a wrongly assessed document. When selecting this menu item, a pop-up window appears and display the list of the last accessed documents, beginning by the last accessed. Icons show the status of the document:

- if the document is validated, a green mark is displayed. If the document is not validated and in the pool, a "highlighting" icon is displayed
- If the document belongs to the pool, a dashed blue box before the name of the document is displayed.
- If the document is relevant (contains highlighted passages), a plus sign is displayed; otherwise, a negative sign is displayed.

An example of the history popup window is shown below.



Under the title "**Collections**" is the list of collections to be assessed. In INEX 2007 (ad hoc task) there is only one such collection, the English Wikipedia collection.

The left or right arrows on the status bar move the focus to the previous or next collection, where there is at least one element to assess (since there is only one collection, no change will occur).

Clicking the hyperlink of “Wikipedia (English)” will take you into the sub-collection view.

4.3. Sub-collection view

The sub-collection views allow you to browse the different sub-collections within the Wikipedia collection. Sub-collections within Wikipedia are based on the alphabetical order, as depicted Figure 3. The first link of the page let you browse the Wikipedia sub-collection starting from "" to "Ali Baba...". This part will then be in turn divided into other sub-collections within the "" to "Ali Baba" range. Eventually, the last sub-collection view will contain a list of Wikipedia documents. Note that this view will show all articles within the collection, and not only those that need to be assessed.

For each possible sub-collection, there is an indication on the number of documents to be assessed in it (if this number is greater than 0), both for documents that were initially in the pool and for documents you chose to assess while browsing in the Wikipedia collection: You are free to assess more documents that there are in the pool, and it is advised to browse to documents that might contain relevant information if you can.

The left or right arrows on the status bar move the focus to the previous or next sub-collection, where there is at least one document to assess. You can also directly click on a link to a sub-collection.

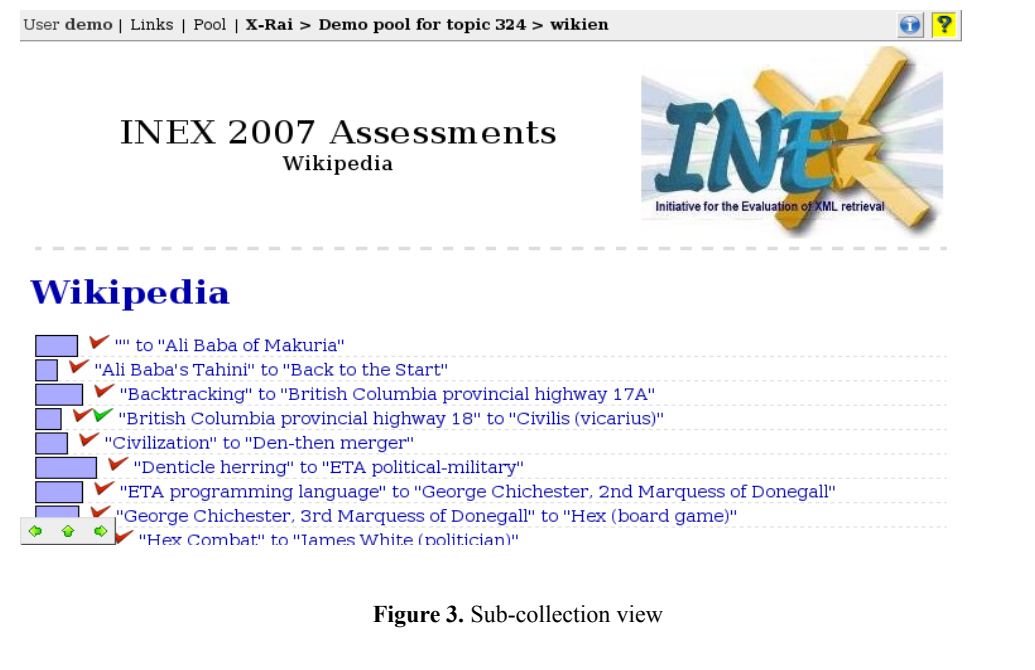


Figure 3. Sub-collection view

4.4. Article view

It is in this article view that elements can be assessed. The article view (see Figure 4) displays all the XML elements of an article together with their content. There are two types of objects within an article view: XML elements and passages. The latter are defined by the assessor while highlighting whereas the former are predefined by the XML file. A highlighted passage in the interface has a yellow background. Note that you should take care of not selecting colours for keyword highlighting too close to the colour X-Rai uses to mark highlighted passages.

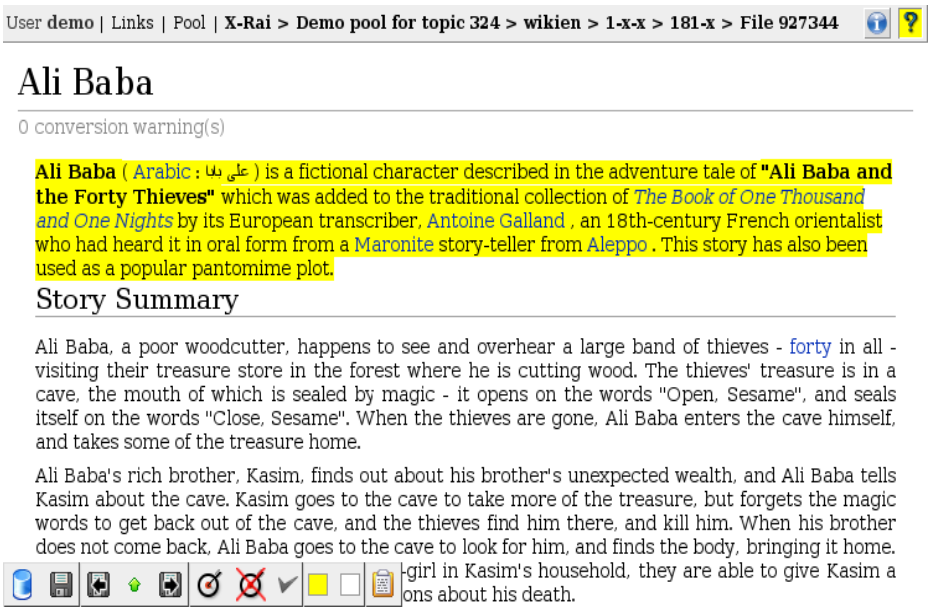


Figure 4. Article view

Highlighting

During the highlight phase, you should identify only relevant (i.e. totally specific) passages by highlighting them. Passages can span over XML element boundaries. The passage limits are predefined by a pre-processing of XML files and correspond “more or less” to sentence boundaries. A consequence of this is that you should highlight the smallest passage that encloses the only relevant information if the predefined boundaries do not correspond exactly to the totally specific fragment. Another consequence is that, it is not necessary to highlight from the first character to the last one – which might be impractical in some case. For example, to highlight the previous sentence, you could start highlighting at the "o" of another and end on the "a" of case.

To highlight a passage, select it with the mouse as you would do in any word processor or text editor, and click on the square with the yellow background (or press “h”).

If you make an error, you can unhighlight it by selecting the non relevant passage and clicking on the square with the white background (or press “u”). If you highlighted too much text, it is easier to unhighlight only the non relevant part.

Note that adjacent passages are merged together: You can highlight large regions of text in more than one step if this is more convenient.

Figure 5. Status bar (article view only)

The disk icon (here disabled): saving your assessments
The disk icon with the left (respectively right) arrow: save (if necessary) and goes to the previous (respectively next) document to assess.
The up arrow allows you to go up to the sub-collection view.

The eye (if applicable): shows or hides the pool elements

The target is used to set the BEP. The stroked target is used to remove the current BEP (if it is already defined for the document).

The mark reflects the status of the document: completely assessed and validated (green), completely assessed but not validated (red), and not completely assessed and not validated (grey). You can validate a document (i.e., mark it as finished) only if the mark is red.

The yellow/white square permits to (un)highlight the selected passage.

The clipboard shows the boundaries of the currently selected passage (as a couple of XPath expressions). This can be useful e.g. to submit bug reports.

Best Entry Point

Focussed structured document retrieval employs the concept of best entry point (BEP), which is intended to provide optimal starting-point from which users can browse to relevant document components. In INEX, you are requested to indicate one and only one BEP for every document that that has relevant content (that has highlighted passages). No BEP should be defined if the document is not relevant (i.e. does not contain any highlighted passage).

To set the BEP within a document (i.e. to be in the BEP mode), click on the  button (or press b) and then click on the position that you want to set as the BEP of that document. It is not possible to set the BEP at an arbitrarily position within the document. The same constraints to those used for highlighting apply for the BEP. In order to help you to know where the BEP will be located, when the mouse pointer is over a Wikipedia text and that you clicked on the "target button", the BEP symbol should appear at the position it would be set if you have clicked. Also note that there are one and one only BEP per relevant document.

Note that although you can set the BEP at any moment, we recommend that you first highlight and then set the BEP.

To remove any previously set BEP, simply click on  (or press shift+b).

Validation of assessments

When you have finished to assess an article, you have to validate it: By validating an article, you guarantee that *every* relevant passage of the displayed file has been highlighted. A BEP has also to be set before you can validate the file, unless the article does not contain any relevant information. In this latter case, setting a BEP would not make sense.

The status of a document is displayed in the status bar when viewing it (or in front of its name in the sub-collection views). The red mark  means the article is not validated, while the green mark  means that it has been validated. You can change its validation status only when viewing it, by pressing the key f or by clicking on the mark. Note that an article is automatically reset to the non validated state whenever you highlight, unhighlight or change the BEP.

Please keep in mind that an article is assessed if and only if it has been validated and saved (see below).

4.6. Saving your assessments

The assessment tool this year does not automatically save the assessments, but you NEED TO SAVE YOUR RELEVANCE ASSESSMENTS by clicking on the disk icon:

The icon is disabled (grey shade) when all assessments are saved.

Be warned that Opera does not provide a way to prevent from exiting a page without saving assessments. PLEASE ONLY USE THE X-RAI INTERFACE TO NAVIGATE INTO THE SITE as this is the only way to prevent you from leaving a page with non-saved assessment(s).

X-Rai reference card

Icon	*Shortcut*	*Action description*
All views within a pool		
	1	Highlight the previous (sub)collection or document to assess.
	2	Go to the container (sub-collection for an article, etc.)
	3	Highlight the next (sub)collection or document to assess
Article view		
	control+s	Save the current assessment
	p	Hide the pool elements
	p	Show the pool elements
	b	Set the BEP
	shift+b	Remove the BEP
shift + 	9	Go to the previous article to assess.
shift + 	0	Go to the next element to assess.
Article view - assessing		
	h	Highlight the currently selected passage.
	u	Unhighlight the currently selected passage
	f	Mark the article as finished
	f	Mark the article as not finished